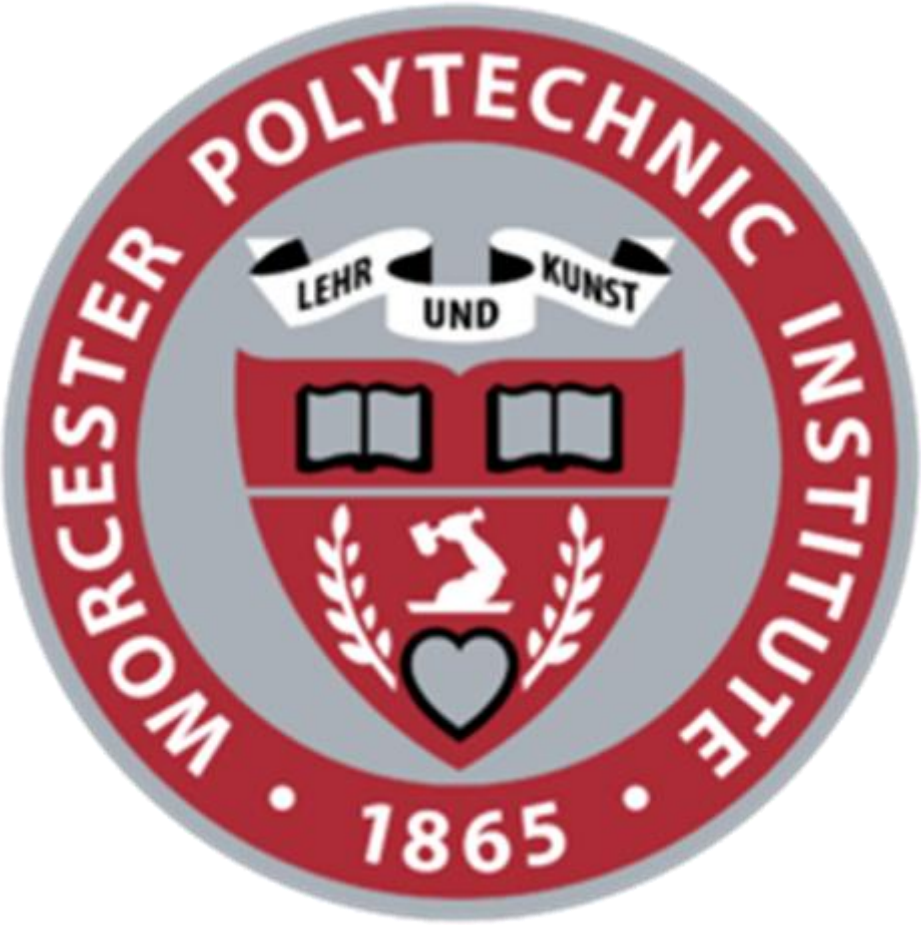
MIS 584: Business Intelligence

Fall 2021

Team 10 Project Final Report

Bhavin Vinod Jain, Brendan Jacques, Matthew Finn, Robert Johnson, Utsav Nandlal Jain



Executive Summary

The client company, Special Tees, is interested in moving beyond the "child" phase of Business Intelligence (BI) maturity, which they are currently at. Thus, our main goal in this report is to assess BI potential at Special Tees. We plan to do this in several ways. First, we will analyse several case studies which are relevant to their current BI situation and where they would like to be. Secondly, we will create two dashboards and a prototype predictive model, similar to those that could be beneficial to Special Tees. Lastly, we will be analysing documentation pertaining to BI implementation. Our first dashboard will analyse customer complaint data with the hopes of understanding customer complaints, their origins, and how effective the sample company is at solving them. Our second dashboard will visualise analytical data related to product sales, with the goal being to demonstrate to our client the information that can be gleaned from existing sales data. And finally, we will create a predictive model that will assess customer retention at the sample company. Ultimately, we hope to demonstrate that these types of models and these case studies, if applied at Special Tees, will be very beneficial to the company.

Having conversed with the client company, we understand that they would like some insights into seasonal staffing, inventory orders, and customer preferences. Thus, we plan to address several of these problem areas in our project. While relevant sample data on seasonal staffing could not be located, our dashboards and predictive model will assist in understanding sales, inventory, and customer preferences. Our first dashboard will look into customer complaints. The goal with the different plots on this dashboard is to answer questions such as "what types of complaints are we getting?", and "what percentage of complaints are we able to resolve?". This is one way that a company can get to know its customers better. Good customer service can boost a customer's loyalty to the company. Similarly, our predictive model looks into customer retention, but does not use a predictor pertaining to customer service calls. Instead, it seeks to solve questions such as "what factors play the biggest role in predicting whether or not a customer will be retained by the company?", and "how accurate is our model given the data provided?". Finally, our second dashboard will demonstrate to the client methods of tracking how different categories of products perform when it comes to profits and individual sales based on the time of year as well as the regions products are being sold in. This would allow the client to make more informed decisions about what their busiest times of year are,

what items to stock more heavily relative to their historic profit, and the preferences of their customers by region over different periods of time.. For Special Tees, these types of models can be applied to any number of business problems they may encounter.

Thus, we will have several deliverables for our report. First, we will critically analyse two case studies. The first of which analyses BI implementation and success at GAP Incorporated, and the second which does the same at the Brazilian electronics retailer Magazine Luiza. We will also have our two operational dashboards and predictive models. These dashboards will be able to answer the business questions provided, and will be fully interactive. The customer complaints dashboard consists of eight plots and will include a slider so that the user will be able to customise the start and end dates of interest for their analysis. The Sales Analytics dashboard consists of four charts, all of which are fully interactive, as well as filters to only show data pertinent to specific years or specific categories/sub-categories of products. Each dashboard will be broken down by plot within their sections, and the files for them will be attached with this report. A discussion of the methods, results, and implications of our predictive model will be discussed at length in the same section of the paper as the dashboards. Some additional information about the model such as some relevant plots of the results can be found in the appendix, as well as a sample of the data used in the modelling process. Following this section of the report, we will discuss ideas for implementing BI solutions at Special Tees, relying heavily on Kotter's eight-step approach. Then, we will discuss some information about data, data collection, and data quality. These three things that are essential to accurate and helpful models and dashboards. Finally, we will conclude with a summary of the results, and some final thoughts and recommendations for Special Tees.

Introduction

This business intelligence project is in support of the client company Special Tees. Special Tees is a screenprinting and embroidery clothing retail store in Kingston, Massachusetts (MA). Its services span over a few counties in the south shore MA area. The business creates personalised apparel and is available for online and in-person ordering. They also have some in-store stock from popular brands like Nike, Champion, and Under Armour. A quote from the company website explains their services well: “We offer: Team Sports, Special Events, School Fundraisers, All Season Wear, Personal Designs and much more.” (Special Tees).

The business has three departments in addition to management which has one member. The retail department has three employees and they deal mostly with the in-store stock. The screenprinting department has four employees and the embroidery department has two employees. The scope of this project deals with data that can affect each department at Special Tees. Some insights that they'd love to obtain some analytics for decision making about are seasonal staffing, inventory orders, and customer preferences.

The status of business intelligence at Special Tees could be categorised in the "child" phase of the BI Maturity Model. For reporting purposes, Special Tees uses about 60% Google Sheets and 40% Microsoft Excel. Sheets is prioritized for data that they'd like to share online within the store. Excel is prioritised for graphing purposes and thus, decision making. This 60/40 split may not necessarily be very well scalable when growing the business, hence there is a need to develop a more structured BI solution.

For other software, Special Tees uses a Point of Sales (POS) system called Microbiz. This is for the sales floor and to manage the customer database. Microbiz can be considered a data warehouse. The company currently does not have a separate BI team to analyse, visualise, and interpret different metrics. Special Tees currently has two main areas of data that they look into for decision making purposes. The first is sales reports. The company gets an idea of the sales metrics, like what % of business came each month and from each department, to help with planning. The second is inventory reports. They use an overview of different departments (in this case for inventory, so jackets, blankets, shirts, etc.) and when they do their ordering they have an idea of which departments are most profitable.

While performing research for this project, we took inspiration and guidance from the following case studies:

1. "Data Analytics and Applications in the Fashion Industry: Six Innovative Cases" by Yue Du

Gap Inc. is a leading US-based global apparel retailer that was founded in 1969 by Donald and Doris Fisher. Gap Inc. currently owns seven brands: Gap, Banana Republic, Old Navy, Athleta, Intermix, Hill City, and Janie and Jack, offering apparel, accessories, and personal care products for men, women, children, and babies.

To understand customers' interests and exploit those insights for fashion trends, the company collects data from Google Analytics, Google Trends, social media, the company's own sales, brand websites (customer browsing and purchasing history, click-through rate, and

time on the website), customer databases (feedback, demographic data, etc.), and even Gap's external vendors.

Google Analytics help track and report a company's website and app real-time data such as number of users, total number of pages viewed, total number of sessions (the period of time a user is actively engaged with the website or app), customer demographic data, the keywords and sites that referred them, and more ("Get to Know Your Customers"). This data can be used by Gap Inc. to better evaluate the performance of its website and app and gain more valuable data about its customers.

The company would always buy a big amount of merchandise for each season, resulting in inventory backlogs and deep discounts. However, with the company's "in-season open program," the final quantity for manufacturing is determined following a short analysis of real-time sales data from a small number of experimental products. The company's capacity to react to the market has improved thanks to data analytics technologies. Company should keep track of who buys what based on age, gender, location, etc. to avoid any inventory backlogs and deep discounting problems which are being faced by them.

2. "Critical Success Factors for Big Data adoption in the virtual retail: Magazine Luiza case study", by Bruno Felix, Elaine Tavares, and Ney Cavalcante.

This case study focused on Magazine Luiza and the programs it was undertaking to apply Big Data analytics to its online retail business. Founded in 1957, Magazine Luiza is one of Brazil's largest retailers in the field of consumer electronics, encompassing over 1000 retail stores and servicing 36 million registered customers. An early pioneer in the field of online retail, the company attributes its incredible success in part to its extensive online shopping website and services, both of which it was seeking to improve upon with data analytics tools. During the period the case study covers, their chief BI tool in development was Project Bob, an AI-powered tool designed to personalise a customer's shopping experience based on their saved browsing and purchase history.

The main reason why this case study was of interest to us was the guidance provided by staff at Magazine Luiza to other companies planning to implement BI solutions. Chiefly, they cautioned that a true implementation of BI tools to their fullest potential will require

rethinking large portions of a company's business model and workplace procedure. For these shifts to be successful, upper management must be strongly involved in the process, clear lines of communication is critical to avoid internal confrontations, and the benefits of both the tools and the data those tools provide should be available company-wide.

The Proposed BI Solution

This section will begin with a description of business intelligence (BI) and then discuss and explain the business intelligence contents in the context of the client company.

Business intelligence is a broad term that covers the technologies, applications, and processes for gathering, cleaning, storing, analysing, summarising, and visualising data to help users make better decisions (MIS 584 Week 1). Business intelligence answers the following questions: What happened? When? Who? How many? These questions can be answered through a variety of analytical processes, including descriptive, predictive, and prescriptive analytics. Business intelligence includes reporting, automated monitoring, headboard creation, scorecards, online analytical processing, and ad hoc queries ((Lutkevich & Burns, 2021)). The BI framework deals with people, processes, governance, and management in order to make a complete decision about the business' needs.

The multitude of components in a standard business intelligence framework all serve a different purpose to meet the specific needs of particular organisations. One component is source or operational systems, also known as transaction processing systems. This is where a business intelligence initiative can have a successful start. Transactional data is created within these databases and is the bulk of the data that is needed for data warehousing. A typical next step is the "ETL" component, which stands for extract, transform, and load. This process gets the clean data into a data warehouse and ready for modelling. Data modelling, the next component, answers the question "what do we need from our data?". It also helps us to better understand the nature of the data, like the format and its relations to other sources of data. This component is extremely beneficial because it helps an organisation become organised and ready for data warehouse building. When the company finally has a data warehouse, they can be ready for their analytical work and data maintenance over the lifetime of their organisation. Other components that are ready to be tackled when the data warehouse is set up include data visualisation, monitoring and alerting, and querying. These components serve analytical teams to become better informed about their operations and their future.

With this structure in mind, we propose first constructing a new data warehouse for our client company that stores transaction data, individual product sale data, customer data, and

customer complaint data. Implementing protocols for collecting the above information is within the company's ability; transactional data is already being collected in spreadsheet form within Google Sheets/Microsoft Excel, and product sale data and customer data can be extrapolated with relative ease from that transaction data. Further, Microbiz has its own API infrastructure, so it's possible for our client to fill in missing data or verify existing data by calling it from Microbiz's servers. The only points of difficulty would be the establishment of a customer-complaint logging system as well as protocols for inputting all the above data into an ETL system for entry into the warehouse.

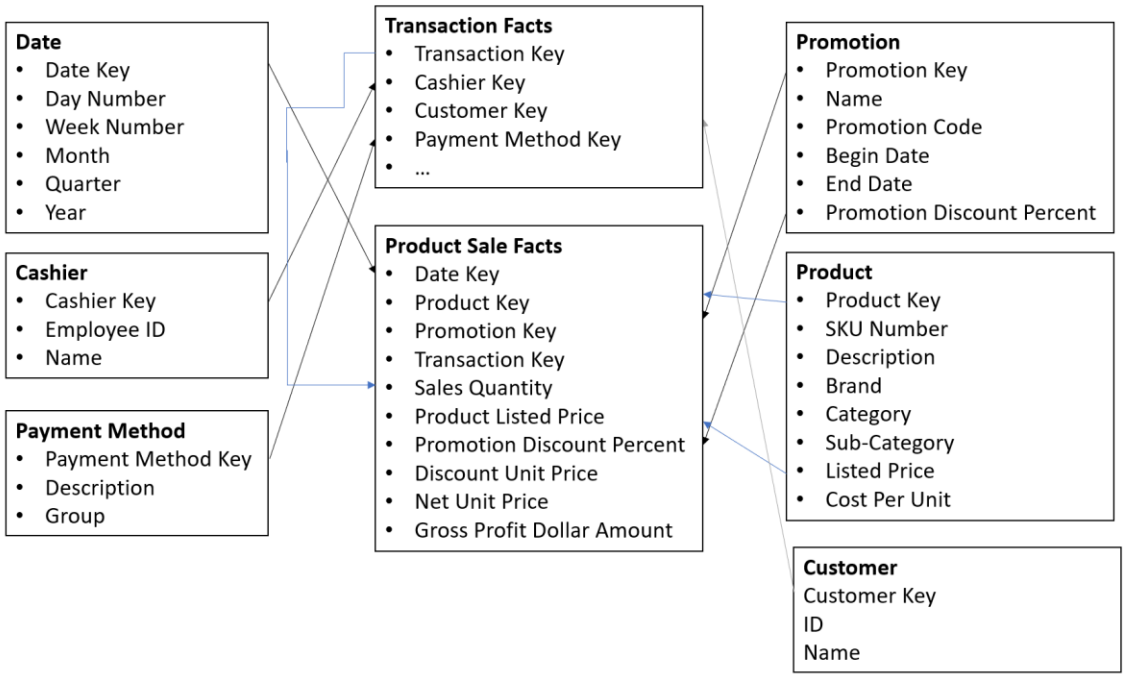


Figure 1: Special Tees' "Product Sale Facts" Schema

Once the data warehouse is established, we would then implement BI tools that can interface with the warehouse to better visualize the data for our company's needs. As it stands, Special Tees does not have a dedicated BI analytics department, so it's important that the tools we provide them can cover a wide range of use cases and require very little programming experience to get functioning in order to get the most lasting returns. To that end, we have three prototypes that we will present in the following sections. The first is a dashboard for visualizing customer complaint information, which we hope will provide prescriptive feedback to the client as to the faults customers see in their services so that they can respond accordingly. Second is a dashboard for visualizing sales analytics, with the goal of giving the client more information as to what products are the most profitable in what regions, what times of year, and under what discount conditions, among others. Third is a

predictive model we have developed in order to determine what factors most contribute to customer retention based on past data, with the goal being to provide our client with the information needed to maintain a consistent customer base in the long-term.

Complaints dashboard

The dashboard depicts the complaints by the customers.

The main idea behind creating this dashboard is to know the type of issues the customers are facing and accordingly allow the company to make their own strategies to combat the issues.

The users of the dashboard are the area managers, supervisors and the higher management.

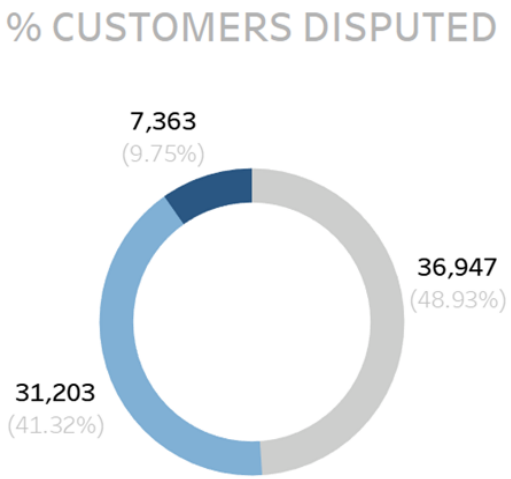
Managerial questions that can be answered are:

1. What is the total number of complaints over time?
2. How much % issues the company is able to resolve on time?
3. What % of issues are ongoing?
4. How much % of customers were disputed v/s the customers who weren't disputed.
5. Which states have the highest number of complaints?
6. What are the most frequent complaints that we're getting?
7. What % of complaints come from different mediums? Which one is the highest?

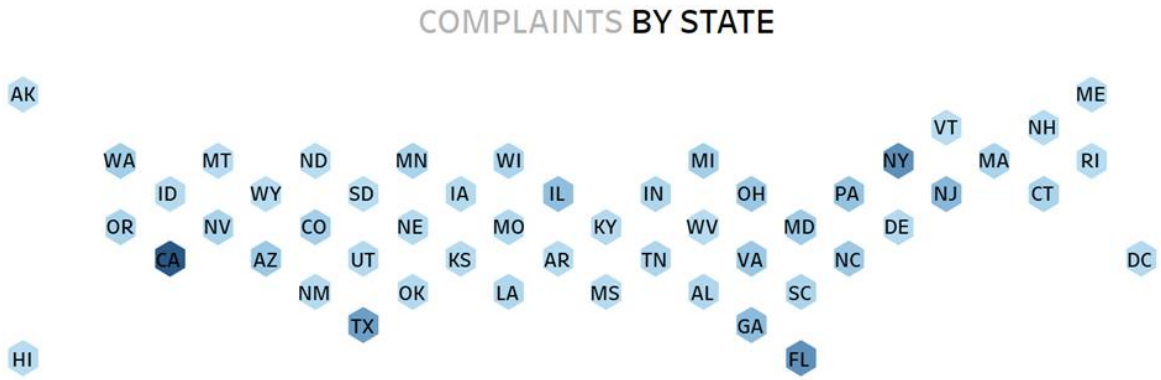
The dashboard has different components:

1. We get to see the total number of complaints over time, which can allow the company to see if there is any trend and also the number of complaints every month.
2. The second component is about the time response %, which tracks how much % of issues the company is able to resolve on time.
3. The In progress bar keeps track of the total number of cases which are ongoing and also gives the % of issues that are ongoing.

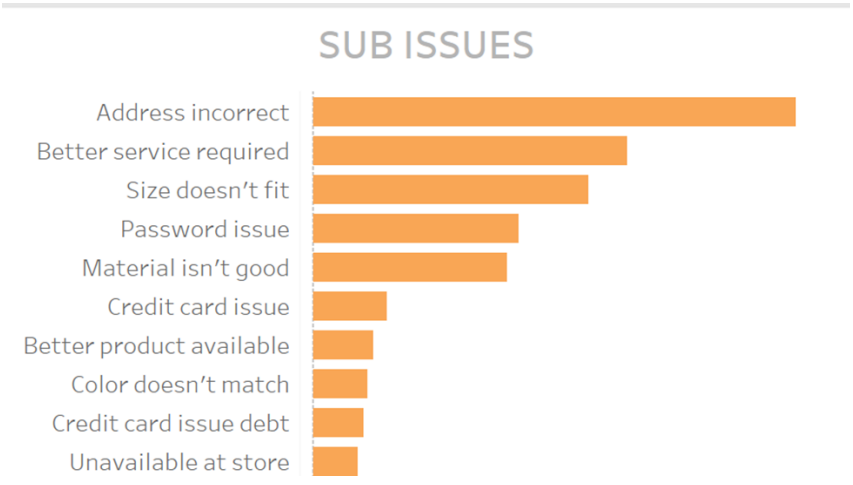
4. The % customers disputed graph gives us the % of how many customers were disputed v/s the customers who weren't disputed.



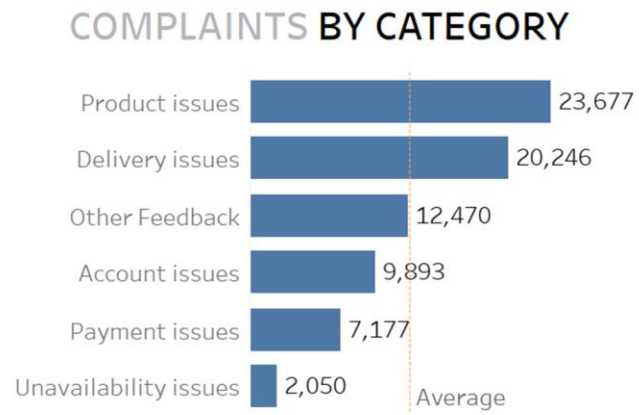
5. The Complaints by State graph gives us the count of total complaints according to each and every state and forms a kind of a heatmap with the lighter colours indicating lower number of complaints while the darker colours indicating higher number of complaints.



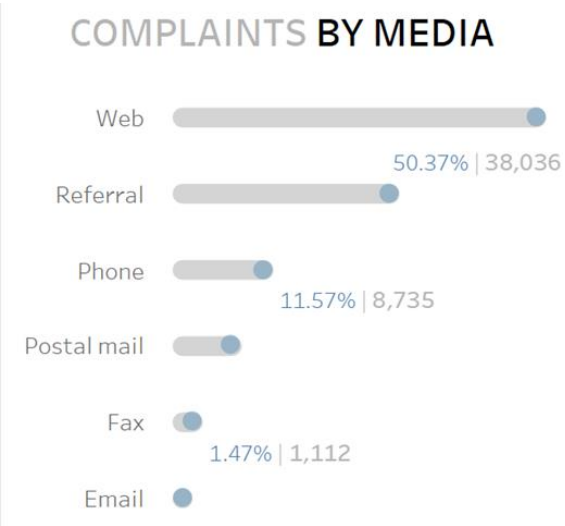
6. The sub issues graph is a bar chart indicating the issues with the highest number to the lowest number. This graph can help the company to know the issues which happen the most and focus more on those.



7. The complaints by category form the higher level view of the type of issues the customers are facing.

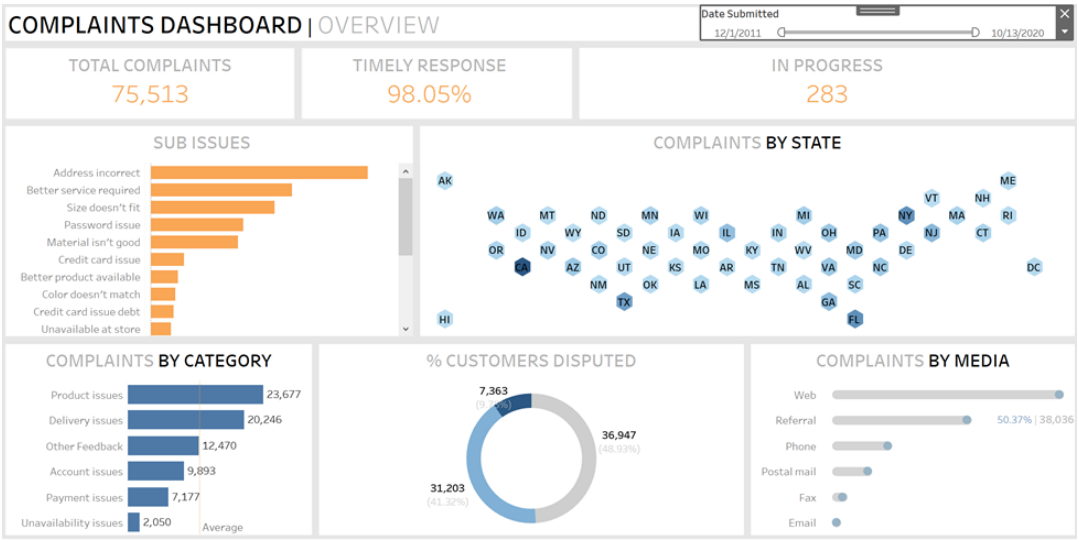


8. Complaints my Media graph gives us the total number of complaints raised from different media such as Web, phone, mail, fax etc. Through the media, the company can analyse if the complaints are mostly due to online sales or store sales.



In the end, we take a filter of different date ranges that can be considered to visualise the issues and the KPIs for those particular days.

The overall dashboard can be helpful as it tracks complaints and helps the company in improving the customer satisfaction score.

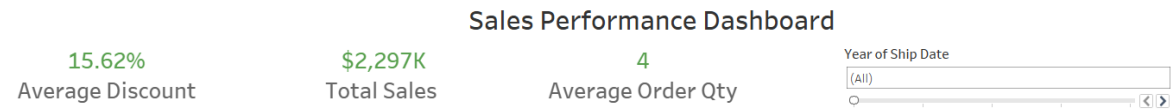


Sales Dashboard

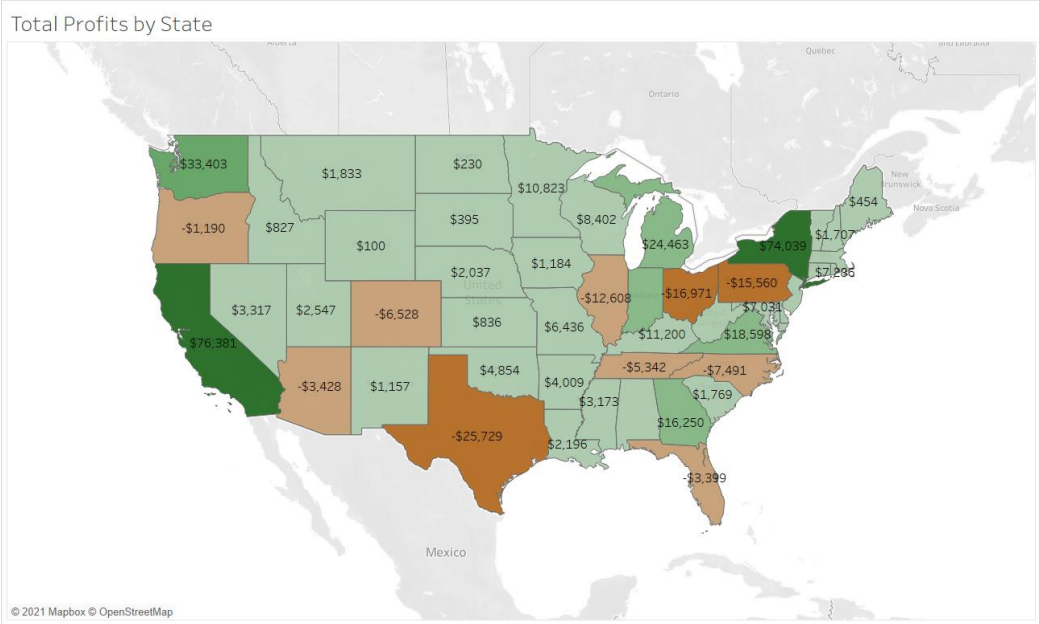
Next, we have our sales analytics dashboard. The goal of this dashboard is to demonstrate to our client company different ways of analysing the sales data across all product types and all serviced locations.

Managerial questions that can be answered are:

1. Which locations have the most sales?
2. What are the top selling and least selling products?
3. What are the products with the highest profit margin?
4. Which are the months with the highest sales?
5. What is the average order quantity?

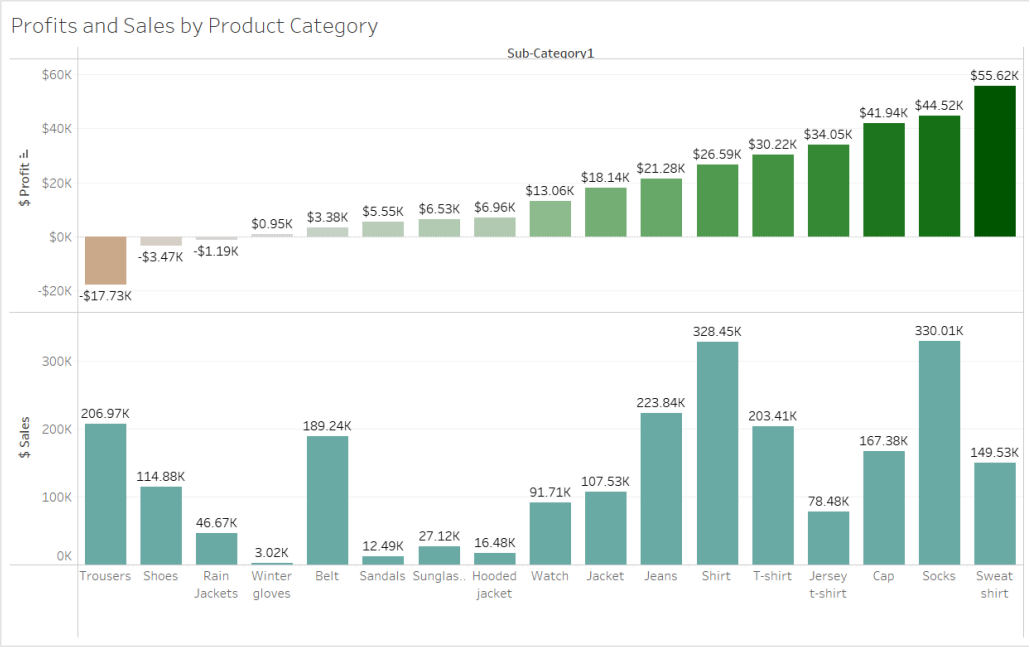


1. The first chart is a map which shows the total profit for stores in different US states. The green shades depict the state-level profit while the brown shades show the loss, with the shade of each colour growing darker depending on the level of the profit/loss. While it's true that our client company only serves the Massachusetts area, this kind of analytics is also possible at a smaller, county-to-county scale and is a good demonstration of the value that can be found by finding the geographic areas where our client's products are pulling in the biggest returns.



2. The second graph displays the Profit and Sales based on Product sub-category. Within the chart are two bar graphs: The bar graph on top shows how profitable each sub-category of item the company sells is, while the bar graph on the bottom shows the total number of sales within each sub-category. Just like with the map chart, the top graph uses a gradient from brown to green to better visualise the degree of profit/loss that is present within the sub-category. We feel this demonstrates a way for our client company to gauge not only how well each sub-category of product they are selling is performing, but how well that sub-category is performing relative to the costs of production. For example, the data in the chart below suggests that while belts earn a mild profit, they are only achieving that profit through sales in the hundreds of thousands, potentially

indicating that belts are being sold for too cheap.

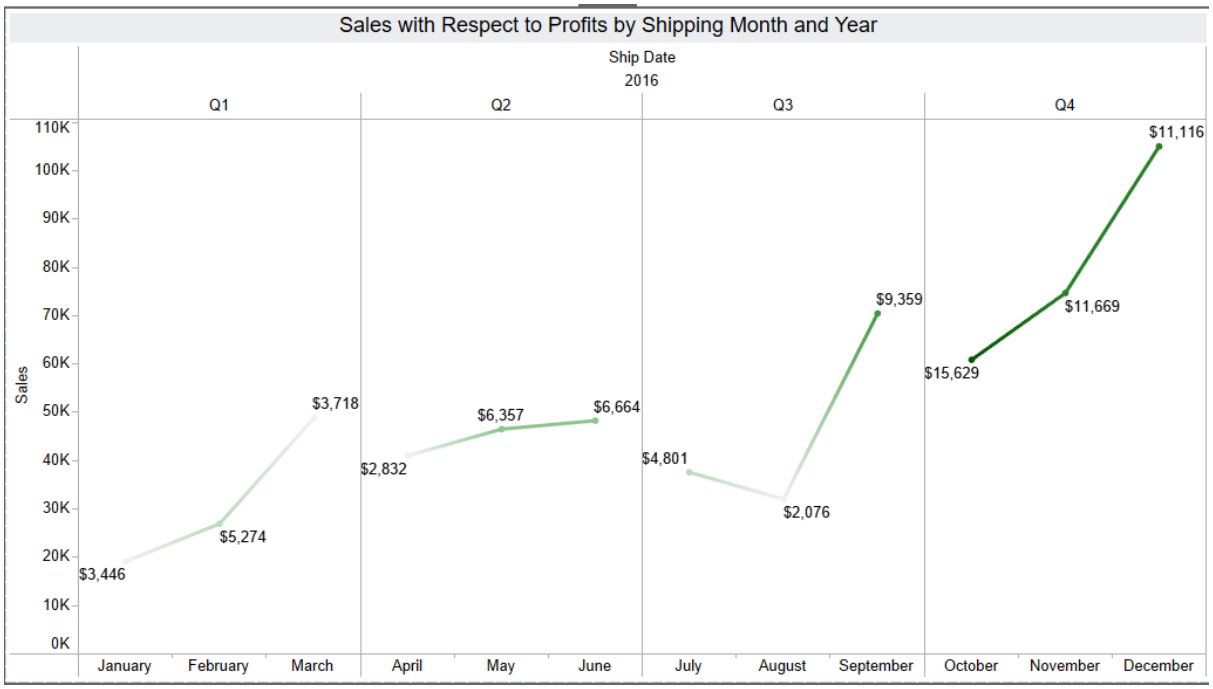


3. Our third chart shows how the number of sales changes with respect to Profits made over differing periods of time. The goal of this chart is to show the user how their products are performing in different periods of time in both raw number of sales and profits earned. By default, the chart displays four graphs separated by year: the x-axis of each graph is that year separated into periods of time in which product sales occurred, and the y-axis tracks the total number of sales that occurred at each time interval in thousands. Just like with Charts 1 and 2, we use a gradient from brown to green on the line to demonstrate the amount of profit being collected as sales increase/decrease over the given time frame. The label at each point in each line graph displays the numerical

amount of profit made at each interval.

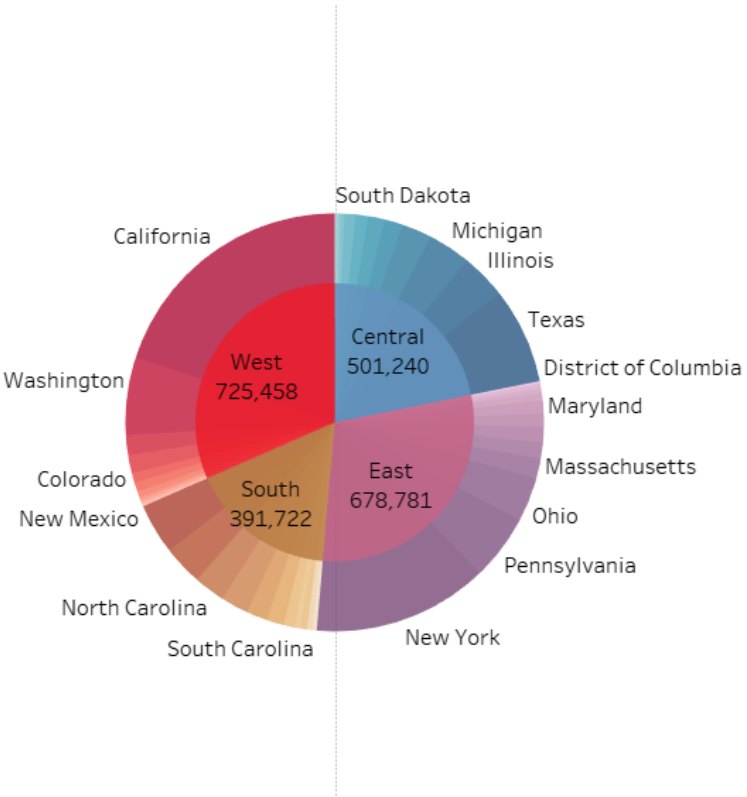
Another feature of this chart is its ability to view its data within time periods at even greater granularity t; by hovering over the x-axis, an option will appear that allows the user to divide the quarters into individual months, then down further into individual days within each month. Each time this happens, the chart will change to show even more charts to visualise sales data within smaller time frames; in the case of shifting to months, it will go from 1 graph for each year (4) to 1 graph for each quarter of each year (16). This does result in the chart becoming harder to read as more charts are introduced, but this can be fixed by the user double clicking a specific graph they want to see in more detail, which will zoom the chart in to more clearly

display that given graph. An example of the chart zoomed in can be found below.

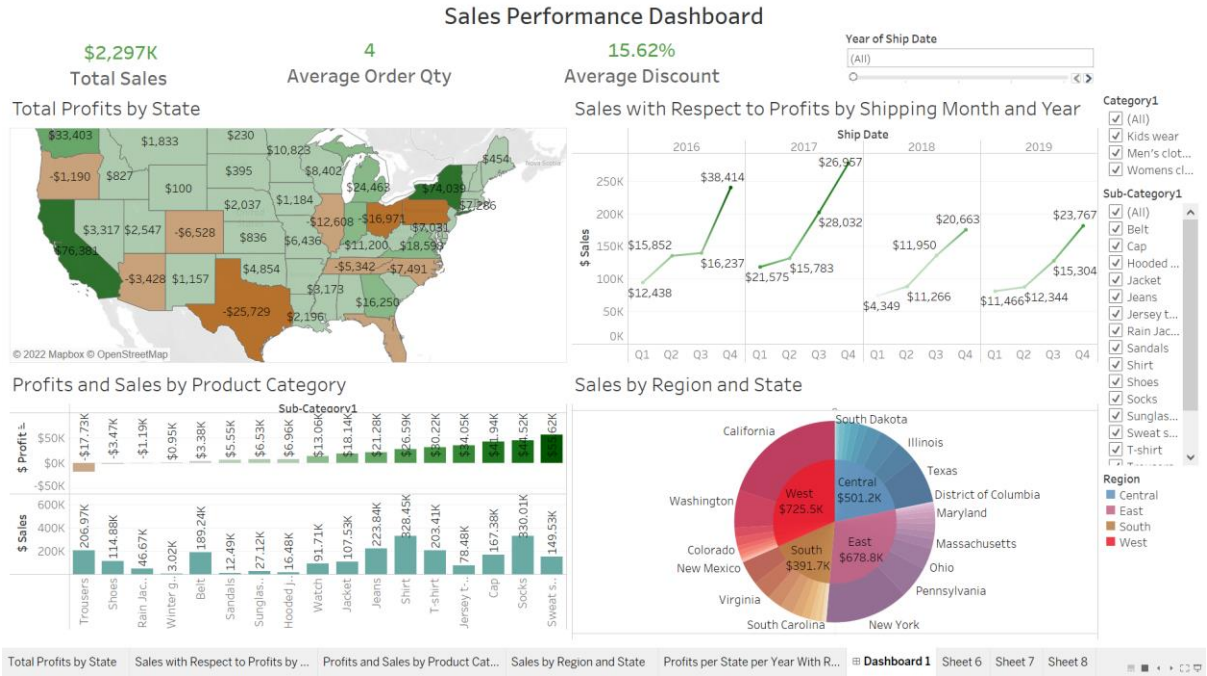


4. Our final chart shows the total number of sales divided into each region of the US, then further subdivided into the States within that region. The goal of this chart is to give further context as to the number of products being sold across the client company's entire service range. This chart will help users to get information about which region has the highest number of sales and which states have the most demand for their products. The number within each region of the chart denotes the total number of sales in each region. The states are sorted within their respective regions from highest sales (darker shades) to lowest sales (lighter shades). Hovering over each section will give

more information about that state in isolation.



Combining all the charts together, here is the final sales analytics dashboard:



On top of the individual advantages of each chart within the dashboard, each one is also interactive and can alter the data being displayed on the other charts. For example, if the user wanted to see only the data pertaining to Michigan, all they would need to do is click Michigan

in Chart 1, and all the other charts would adjust to only show data relevant to transactions that occurred in Michigan. Other examples include clicking on a sub-category in Chart 2 to show only the sales of a specific product, clicking on a region in Chart 4 to only show sales within that region, or clicking a specific quarter in Chart 3 to show sales within only that timeframe. These filters can also be combined for even greater granularity (ie. Showing only sales data in Texas within Q2 of 2019). Lastly, to the side of the dashboard are further filters that allow the user to only show on the dashboard data within specific Shipping Years, product Categories, and product Sub-Categories. We hope these controls will demonstrate to our client company the value that can be attained by modelling their Sales data using Data Analytics tools.

Prototype 3

For the project team's third prototype, a best subsets regression technique was performed. Best subsets regression tests every possible combination of the predictor variables in a dataset. Then, according to the particular statistical criterion of the user's choice, the algorithm selects the best model. The dataset used for this technique consisted of customer retention data from kaggle (See Appendix A). The team believed this dataset to be appropriate for this analysis because the regression technique is able to consider which variables are most important for customer retention. The dataset itself is also fit for a retail store and has similar features to what Special Tees might be considering (average order, order frequency, emails opened, etc.).

The first step to running this technique on the data is to decide which variables should be considered to affect "retention", our target. The team cleaned up the data by removing some unhelpful features and loaded the dataset into R, the software of choice for this analysis. The R package "leaps" has a built-in function for computing best subsets regression that the team used. The output below reports the best set of variables for each model size.

What this means is, for example, if you were to look at each row #2 (there are 3 of them because many features have been used in this analysis) and only consider the features with an asterisk, you would come to the conclusion that the best 2-variable model contains "Year_Created_2014" and "esent". The next step in the process is to compute some statistics to determine which model is the most effective for predicting retention (is it a 1-variable model, 2-variable, and so on). The statistics that the team computed to come to this conclusion were the following: Adjusted R^2, CP (Mallow's CP estimator), and BIC (Bayesian Information Criterion). The output below clearly tells us that the model with 5 variables is the best for

predicting retention. See Appendix B for more statistical plots that helped the team determine the best model for this dataset's regression.

```
Adj.R2 CP BIC
     5  5   5
```

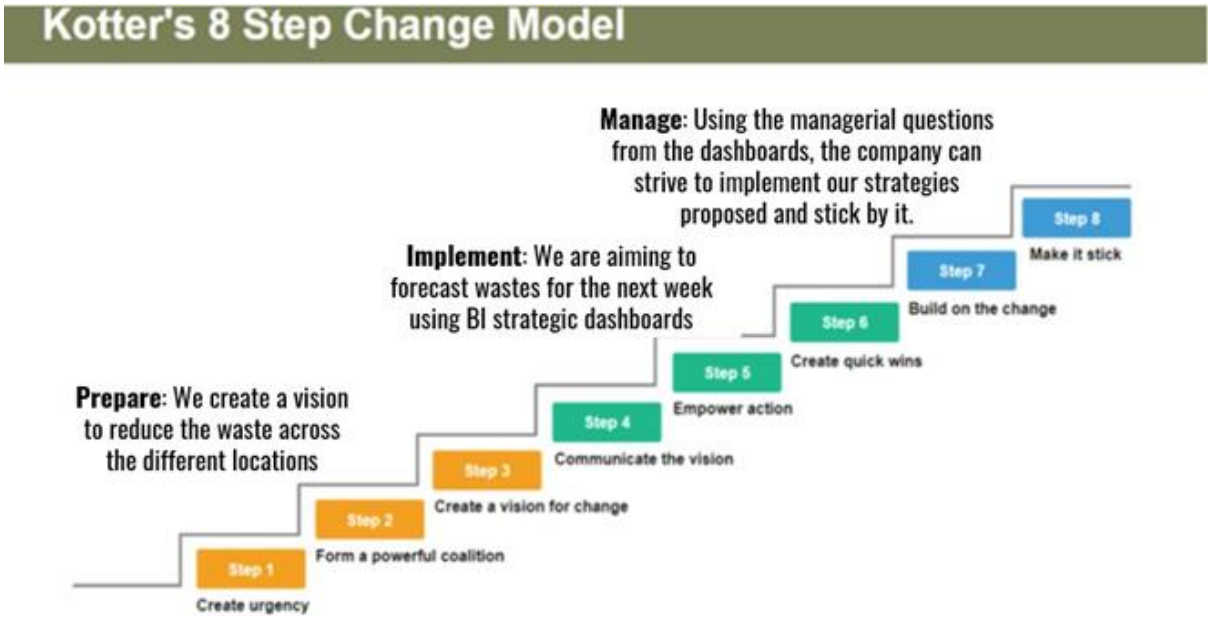
Now, the team is able to conclude that the following variables, together, are the best predictors of customer retention: "Year_Created_2014", "Year_Created_2018", "esent", "eopenrate", "paperless". Some insights can be made from observing these variables. First of all, we can note that when a customer joined either in 2014 or 2018, something positively affected their retention. This could be anything, such as an appealing promotion, effective employees, or high quality product releases. What is important is that this particular company needs to analyse what they did right in these two years and what they can do in the future, similar to 2014 and 2018, to get these customers to stay. Next we consider "esent" and "eopenrate". Simply put, if a customer signs up for email notifications, and especially if they are opening these emails, they are more likely to stay connected to the business and be retained. Finally we have "paperless" which simply means that they signed up for paperless billing, news, and notifications. The team concluded that this could be a feature affecting retention because it means the customer went into their account and made a change, simply meaning that they were active on the website or on the phone.

The potential uses of a best subsets regression technique on Special Tees data are vast. The algorithm itself is quite simple to perform, requiring very few lines of code. Special Tees keeps customer data in a warehouse and it would be feasible to make some insights about their customers if they can choose a categorical variable to target. Retention is just one option, but Special Tees may want to look at in-store vs online sales, the nature of large vs small orders, compare popular categories of inventory (jackets vs blankets, for example) and see what is affecting those sales, and much more. Special Tees already takes a look at sales reports and inventory reports for some soft/objective analysis, so best subsets regression might be an easy implementation for the business and aid in decision making.

IMPLEMENTATION

Without a solid implementation strategy, the BI solution, which includes the data warehouse we built, would perish since there are a slew of issues that can hinder everyone from top management to front-line staff. For example, senior management may have miscalculated the project's scale and consequences, or staff may be resistant to change because they are too accustomed to the system they have been using for decades. On the

managerial side, Kotter outlined a process that can assist businesses in avoiding failure during change, and we give it our highest recommendation. Throughout the program, Special Tees follows Kotter's eight-step approach as a guideline.



Kotter's Eight Steps for Special Tees

Kotter's Eight Steps

According to the framework of Kotter's eight steps to transforming your organisation The first phase Special Tees will confront is to be modest; do not think the company has always been doing well, avoid locking down in the comfort zone without leap forward, and **'establishing a sense of urgency is essential.'** We should highlight the number that can influence all the stakeholders, such as a statistic that shows that there are sales in the summer months, the fact may not be visible, but the number cannot lie. Frequent communication, which is also an important aspect of this phase, uses a variety of channels, including meetings, emails, and speeches, to tell all stakeholders about the major problem and provide a vision or plan that the organisation may follow. Apart from discussing the company's benefits, we need to devise a strategy to pique participants' interest and foster an exciting work atmosphere in order to meet the future changes.

While everyone in Special Tees knows the urgency, **'form a powerful guiding coalition'** to be able to push the project forward and under control. Bring relevant people together in a team that includes not only top management but also middle management, supervisors, and technical experts to commit to successfully completing the BI project. As

for the project itself, Special Tees should assess its importance and allocate business resources to assist it in terms of people, budget, equipment, facilities, or anything else that can aid the transformation process. What is the goal of the company? What has to be changed? What steps can we take to make the vision a reality? The advancement of Special Tees will be more likely to flourish in the future if these questions can be answered. "... without a credible path out of a poor position, individuals and organisations have a remarkable aptitude for disregarding worrisome evidence and continuing to act in established patterns," according to The benefits of data warehousing: why certain businesses enjoy exceptional payoffs.

Furthermore, we recommend that Special Tees '**establish a change vision**' on the topic of profitability in several areas. The vision can be linked to stakeholders' interests and the well-known knowledge quantity of all employees in the organisation for management purposes.

The information that underpins the vision is much more important than the vision itself. Similar to phase one, 'establish a sense of urgency,**' 'communication the vision'** allows the project guiding team to send out a strong message to explain potential benefits why the transformation has been related to everyone in the Special Tees.

Everything is just theory, before taking tangible action. **'Empower others to act on the vision'** notice the company to remove all the obstacles and take action. At this phase, some of the higher management may still cast doubt on the vision or the system we want to implement. In this way, as the project guiding team, we should detailly explain the KPIs we have created and what kind of managerial questions the BI system can answer. Finally, make them affirm that implementing the system will help business growth in the long-run.

'Plan for and create short-term wins' phase, Special Tees should achieve some benefits from implementing the BI system. Such as the percentage decrease in complaints every week, and the percentage increase in revenue every week. The result will be the best cure that can convince stakeholders that we will succeed in the future.

Now, we understand that short-term wins are essential for successful change because they provide credibility, resources, and motivation. In the **'consolidate improvements and produce still more change'** phase, keep up these good works we have done, and continue to improve the changes we have made. For example, assuming the BI system and data warehouse has successfully been implemented in Special Tees. However, we can still find

out drawbacks to the system. The data we have collected may not be precise, but creating new KPIs fits the company goal, a new visualisation model that can support decision making in the Special Tees.

Final phase 'Institutionalise the new approaches', The best result of the BI solution is part of the organisation strategy, this strategy can change over time, while the BI solution continues to iterate and meet the company's needs.

Data Quality

Data quality refers to both the attributes associated with high-quality data and the techniques used to test or enhance data quality from a technological standpoint. Before we begin to evaluate, we must first ensure that the data is of high quality. We need to improve the quality of data by cleaning the dataset, removing data such as null values and unstructured values, and refining the columns that we need for our analysis. We cleaned the data in Excel, but as long as Special Tees develops a Data Warehouse, all important data will be collected by ETL tools, as well as cleaned and restructured on SQL, because we employ a large dataset that spans more than four years.

Most data scientists/analysts nowadays spend a significant amount of time cleansing data. Data governance can be a method for ensuring high-quality data while saving time and money. Data governance system ensures that our retail store data stays relevant, up-to-date, and consistent across the multitude of sales and marketing platforms.

Data integrity may be ensured, and unstructured data can be avoided, using this method. Second, establish regulations to standardise data formats that can be used more effectively with the current BI system and help us achieve our aim. Finally, to increase data quality, Special Tees should acquire people and build a data management team, or partner with a third-party expert. Simultaneously, leverage the BI solution to assist them in making decisions that are consistent with the company's goal.

Summary and Conclusion

Herein we discussed what BI is and gave several examples of how it can benefit a small business such as Special Tees. It is clear that BI can, if implemented properly, further the business. Thus, we discussed a case study pertaining to GAP Incorporated. In it, we can see how BI is utilised at a clothing retail company and how beneficial it can be, albeit in a much larger company than Special Tees is. From the GAP case study, we see that they use

internal data such as sales data and data on customer feedback, as well as external data from Google Analytics. One of the benefits of BI is that it is very versatile. Yes, GAP is much larger than Special Tees and what works for them is not necessarily what will work here, but there are very important lessons to be learned from their example. They studied their needs and implemented what works well for them. For Special Tees, perhaps just using either internal or external data would work best, or, if needed, some blend of both may be the better option. It is ultimately up to the management at Special Tees to determine exactly what they need to further the business, and how best to implement it. The dashboards and predictive models demonstrated herein utilise sample data that acts as internal data (sales figures, retention information, customer complaints, etc.). This same analysis could be conducted on external data to better adapt to a changing market, as GAP has done so successfully. But, as has been mentioned previously, BI should be tailored to the specific needs of the business.

Our three BI models created for this report span large swathes of data, and all demonstrate some of the unique capabilities that BI might be able to offer Special Tees. Our first dashboard examines customer complaints. Of course, the goal here is to get a better understanding of the issues customers are having, and enable the management at Special Tees to create strategies to redress these issues appropriately. In the sample data used to create this dashboard, we see that two of the biggest sub-issues facing customers are in regards to incorrect addresses and improper sizing of the products. These issues are relatively easy to remedy for the business, whereas others may not be. Dashboards such as this one give businesses valuable insight into day-to-day operations, and can help discover both the positives and negatives of the current track the business is on in ways not previously possible. Similarly, our second dashboard analyzes sales data across all product types and all serviced locations. This way, we can gain insights into things such as what customers in each state tend to buy, and how much they spend. We can also see what types of clothing generate profits or losses in ways that would be much harder to do without these tools. Likewise, we also created a prototype Best Subsets Regression model to demonstrate another form of data analytics. This model predicts customer retention through several given predictors. This model revealed insights into customer engagement with the business, and suggested that the year the customer joined is important. Then, knowing this, the business can evaluate these results and contextualise them. This type of model, along with others such as decision trees, cost-benefit analyses, and so on help to empower better decision making.

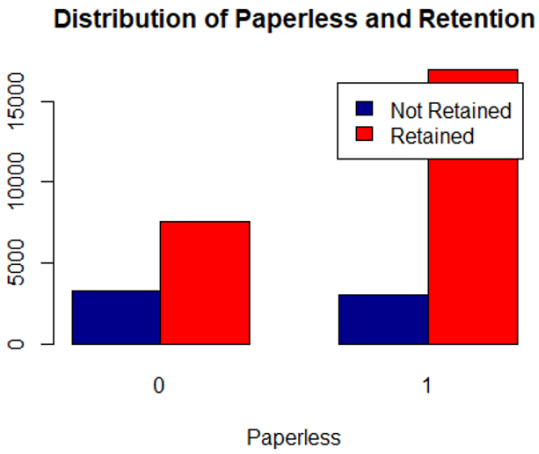
Now, where does Special Tees go from here? We believe that the first steps are to determine what the end goals of BI at Special Tees are and what form they should take. As it stands, we understand that you would like to analyse seasonal staffing, inventory orders, and customer preferences and gain insights into them. So, it will be important to meet with everyone in the business and come up with a plan. It will be important to decide what data will need to be collected or measured and what metrics and methods will be used to analyse it. As it usually goes, the more data, the better, but, as always, make sure it is consistent with privacy laws and guidelines. And, as we discussed herein, keep Kotter's 8 step change model in mind when developing a plan. Communication is crucial, and getting insight from everyone relevant to the project is key. Evaluating progress along the way will also be essential, so decide on some benchmarks or other metrics to ensure things are going smoothly. Finally, make sure you have a backup plan, or some way to act accordingly if necessary. These projects can add a lot of value to a business, so now is the time to begin planning.

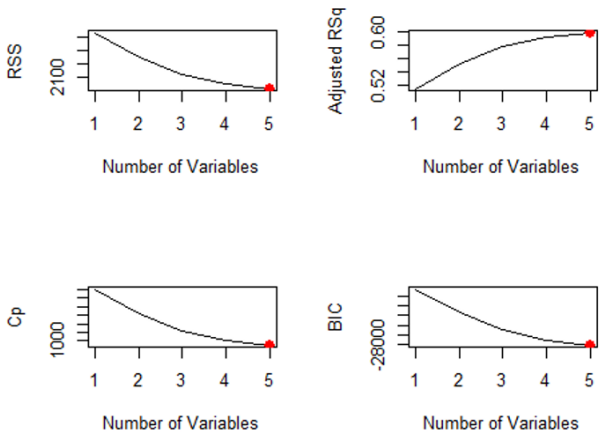
Appendix

Appendix A: Customer Retention Retail Data from https://www.kaggle.com/uttamp/store-data

custid	retained	Year_Create	esent	eopenrate	eclickrate	avgorder	ordfreq	paperless	refill	doorstep
6H6T6N	0	2012	29	100	3.44827586	14.52	0	0	0	0
APCENR	1	2010	95	92.631579	10.5263158	83.69	0.18164063	1	1	1
7UP6MS	0	2010	0	0	0	33.58	0.05990783	0	0	0
7ZEW8G	0	2010	0	0	0	54.96	0	0	0	0
8V726M	1	2010	30	90	13.3333333	111.91	0.00884956	0	0	0
2B6B83	1	2008	46	80.4347826	15.2173913	175.1	0.14117647	1	1	0
99XGVM	1	2011	60	43.3333333	6.66666667	116.55	0.125	0	1	0
U3MP5L	1	2009	64	28.125	15.625	68.1	0.04009434	1	1	0
ELKAGQ	1	2010	45	0	0	46.6	0	0	0	0
3SBQP2	1	2010	34	94.1176471	8.82352941	66.07	0.13378906	1	0	0
YMALVV	1	2010	48	83.3333333	2.08333333	89.04	0.08665511	0	0	1
GW8NT7	1	2009	40	87.5	45	88.3	0.08847737	0	0	0
TFKLD4	1	2009	69	2.89855073	2.89855073	126.81	0.05943153	1	1	1
RAQ92T	1	2010	38	36.8421053	23.6842105	116.77	0.07894737	0	0	0
DMJERJ	0	2011	16	12.5	0	59.25	0.12935323	1	1	0
FJKD8W	1	2011	45	60	15.5555556	49.73	0.04424779	1	1	0
HEPCW7	0	2010	13	61.5384615	15.3846154	156.54	0.11911357	1	0	0
2JZAE3	1	2010	48	8.33333333	2.08333333	101.06	0.03780069	0	0	0
BYS2DU	1	2011	57	17.5438597	0	14.02	0.06600985	0	0	0

Appendix B: Additional Plots from Best Subsets Regression





References

Special Tees. (n.d.). Welcome to special tees. Special Tees. Retrieved December 9, 2021, from https://specialteeskingston.com/.

MIS 584 Week 1, MIS584_Week 1.pptx

Lutkevich, B., & Burns, E. (2021, October 12). *What is business analytics?* SearchBusinessAnalytics. Retrieved December 9, 2021, from https://searchbusinessanalytics.techtarget.com/definition/business-analytics-BA.

Business Intelligence Architecture - what, why, and how. Jigsaw Academy. (2020, June 2). Retrieved December 9, 2021, from https://www.jigsawacademy.com/business-intelligence-architecture/.