Business Intelligence solution for Tatnuck Bookseller

MIS 584-Spring 2018

GROUP 2   **Submitted on: 04/25/2018**

Shruti Agrawal

Rushikesh Naidu

Matthew Weiss

Brandon Werner

Anuj Vyas

Executive Summary:

Goals

Tatnuck bookstore is largest independent store located in Westborough, Massachusetts. The company uses various tools internally for their operational activities such as imerchant and excel. The company uses domain knowledge of general managers for decision making. The goal of the project is to propose a Business Intelligence solution for Tatnuck. This will support the company in near real time monitoring of the day-to day operation, decision making and predictive analysis.

Challenges faced

While trying to implement a comprehensive BI solution for Tatnuck bookseller we faced few challenges which we have mentioned below:

The main challenge we faced is to find the sales data for the bookstore. Different set of metrics are important for different users. To overcome this, we used gift store data and bestsellers data from New York Times.

The second challenge for any effective BI solution is, finding a clean data set that is well integrated between different departments and data sources. For the project we made use of a simulated data set but, in real applications it is not an easy task to find the correct format of dataset which will satisfy the requirements of the BI solution. When identifying the outcomes, as noted above, it is important to understand the available data for the analysis at the same time.

Summary of Project Outputs

When we studied Tatnuck's market position of industry and the solutions being used at present, a BI solution will deliver the clear visibility of the flow of inventory. We have built 2 prototype dashboards which will support the company in managing their operational activities and monitor New York Times best sellers. The various graphs within the dashboards are linked to each other, which would change as the filters are changed or selections are made of a section. If the data is available to the tool online, it can be presented anywhere anytime and will be as real-time as the updates made in the database the returns on investment for this project will gradually increase with time.

Introduction

Company Background

Tatnuck bookstore is mom and pop bookstore with three departments as books, gifts and café. Company has the workforce of 30 employees between bookstore, café and backroom. Tatnuck conducts various workshops and networking events. Tatnuck focuses mainly on employees and relies very little on technology and collects sales, quantity, costs and inventory. For the scope of this project we will be focusing on delivering a solution in the books, gifts and café business unit which covers the operations.

Current Status

Tatnuck uses spreadsheets for reporting and imerchant for monthly sales reports. They do not have any data warehouse and BI team. They use variety of disconnected systems, for example, daily sales are directly entered in Excel each morning while monthly reports are generated from imerchant. They use excel for their daily reports and enter all the data manually. The reporting produced by the company is still heavily dependent on MS Excel spreadsheets and ad-hoc reports. They do not have any data warehouse and BI team. They use variety of disconnected systems, for example, daily sales are directly entered in Excel each morning while monthly reports are generated from imerchant. Currently, the company uses imerchant as the most centralized system and all the data from imerchant is stored onsite. The manager of the bookstore is happy with their current system and thinks implementing BI solution will be good for a big company but not tatnuck. They have an opinion that upgrading to a new system can make things worse.

Proposed Solution

We believe that the Tatnuck bookstore needs a visually appealing solution which will help Tatnuck monitor books on the New York Times (NYT) best seller list. Our dashboard solution will also help Tatnuck track wasted items in their cafe and identify best-selling items in their gift store. We propose a BI solution using Tableau to achieve outcomes.

Business Intelligence

What is BI?

Business Intelligence (BI) is a broad term that covers the technology, applications, and processes for gathering, storing, accessing, analyzing data to help users make better decisions. More organizations than ever before are integrating business intelligence into their daily operations to help them gain situational awareness and improve their operational effectiveness. Organizations are using business intelligence to identify potential sales opportunities for their sales team, or to recognize anomalous customer purchase behaviors. In layman's terms, organizations are using BI to extract value from their data to help them make smarter decisions and become more efficient.

How does BI work?

BI uses visualizations to communicate data to people. It is extremely difficult to look at a spreadsheet and gather any meaningful insights without spending countless hours. By using the visualizations that are typically incorporated into BI systems, users can gather insights by using features like clusters. Developing a BI system is a massive undertaking though and it is no simple task. A typical BI implementation takes at least six months and requires a team of highly skilled professionals who knows what they are doing.
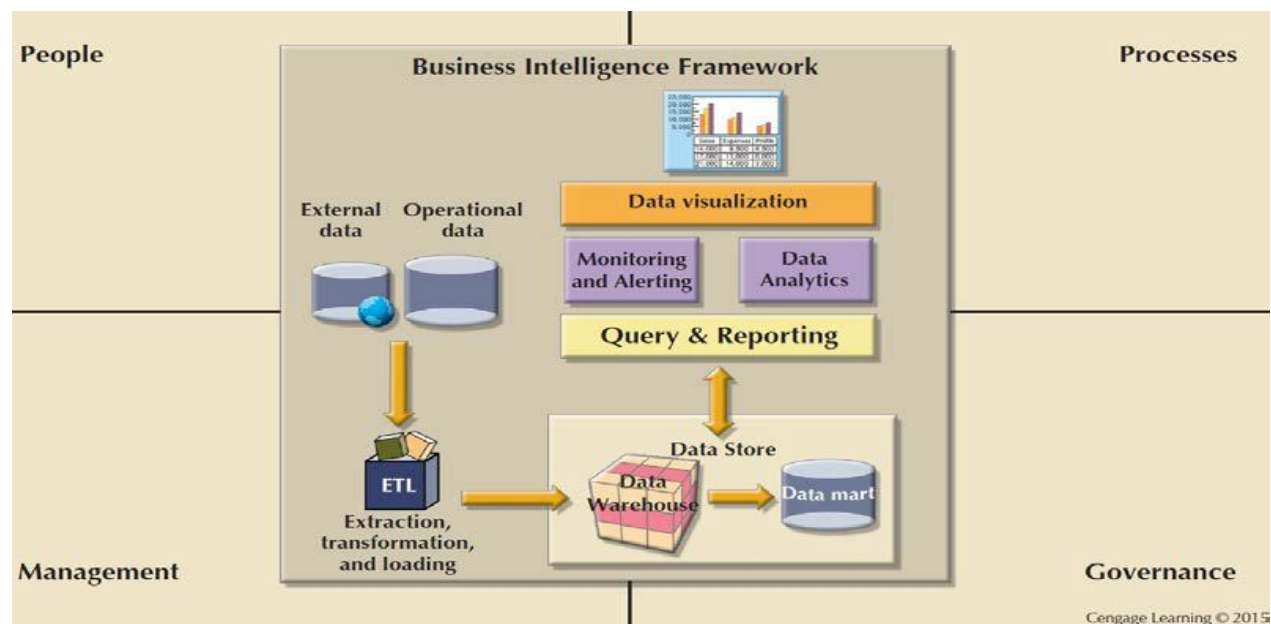

Figure 1: Business Intelligence Framework

Source: (Rob & Coronel, 2009, pp. 517)

Organizations use the Extraction, transformation, and loading (ETL) process to take their data from their operational database and external data from outside sources and implement it into their data warehouse. ETL includes the data enrichment processing, which is comprised of data standardization, data normalization, data cleanup, etc. This is the most resource intensive stage of BI as it is crucial that an organization thoroughly clean their data before building out their data warehouse. Some organizations will also create data marts which are small subsets of data warehouse that provide decision support for a small team of people.

After using ETL to place data into the data warehouse and data marts, organizations use data visualization and data analytics display their data in easy to understand ways that makes it possible for the human eye to pick up on patterns that would have previously gone unnoticed in a spreadsheet or database. This is where everyday users see the value of the BI system and they run queries and generate reports from their data mart to find new insights and conclusions about the data they have. Monitoring is also a key part of any operational dashboard as it can help you gain awareness of what is going on in your business at any time.

What are some BI tools?

Figure 2: Gartner Magic Quadrant for Data Visualization Tools

Source: Gartner (February 2017)

Some of the more popular BI tools are Tableau, Microsoft, and QlikView. There are also other tools like IBM SPSS, SAS, and Alteryx that don not create visually appealing visualizations but are helpful in processing the data. According to Gartner, an industry leader in IT research, Microsoft, and Tableau are neck and neck in the leader's category, while Qlik is a distant third place. Then we have SAP, SAS, IBM, Salesforce, which fall into the visionary category. Lastly, we have the niche players which includes Oracle, Alteryx, Domo, Logi Analytics. It is important to note that many of these tools, particularly in the niche category, serve a specific purpose. Even Tableau, which is considered the market leader, is designed primarily for data visualization rather than data processing (Gartner, 2017). Alteryx on the other hand is one of the best tools for data preprocessing [8]. In fact, many organizations use multiple BI tools together, particularly Alteryx and Tableau.

Who is using BI?

Everyone benefits using BI whether it's the individual team contributor or the Senior Level Manager. We are producing a larger amount of data than ever before and BI tools like Tableau help us identify potential patterns. Even Fortune 500 companies like Exxon Mobil and Kimberly-Clark are using BI systems including Tableau [7].

Benefits of BI

BI tools are relatively easy to use, and no coding experience is required to use them. They are drag and drop which makes them easy to learn even for the non-technical user. They usually have some template dashboard to use which helps when a user is first learning the tool. You can produce nearly any type of chart imaginable with BI whether it's a box plot, bar chart, etc. BI allows you to track your Key Performance Indicators (KPIs). KPIs are an efficient way for teams to track performance, which is crucial for Customer Service, Sales, and Marketing Teams.

Drawbacks of BI

The drawbacks of BI are that it does require the data to be properly cleaned beforehand. If you put low quality data into the BI solution you will get a poor-quality visualization that will not be helpful in any way. Also, BI solutions are expensive especially when you are paying per a user license. Some tools like Tableau are $500 per license, and when you are operating in a large corporation that cost can add up.

As we previously mentioned the massive undertaking and effort required to implement a BI system, if this is the first time your organization is implementing BI, then your organization will need to hire at minimum one or two BI consultants to help with the development and rollout. Ideally your organization will hire a team of experienced BI professionals to bring on staff in-house who have experiencing implementing BI at companies similar to your organization's caliber.

BI Maturity Model

We decided that Tatnuck falls into the Infant stage of the BI maturity model since they do have some static spreadsheet using Excel, but they do not have a data warehouse with all of their data in one place. They are a very manual operation and based on our interactions with them they prefer to keep things the way they are. Essentially with the Excel spreadsheets Tatnuck is using a spread-mart which put them into the Infant stage. However, from time to time, they do produce some ad-hoc reports using Excel, but we believe since they are not Analytical Systems that they do not fall into the Child stage.
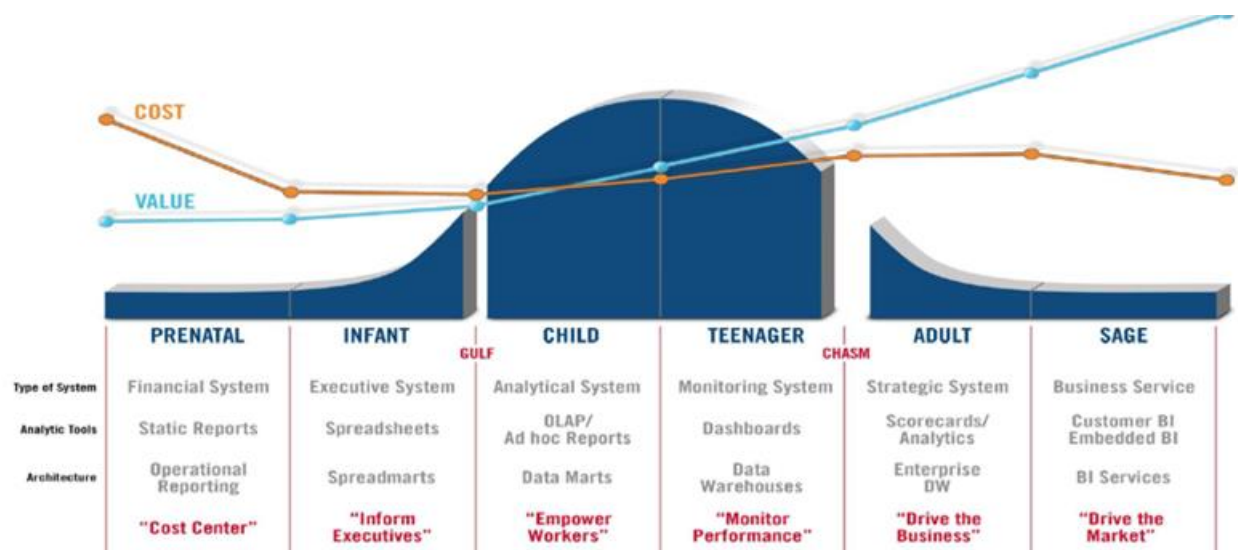

Figure 3: BI Maturity Model

Source: CSG PRO (November 2015)

Proposal

The main components of a BI solution are people, processes, management, and governance. Our BI solution will address all four of these areas and identify opportunities for Tatnuck to improve their operations.

Data extraction, transformation, and loading (ETL) tools	In ETL, we will collect, filter, and integrate operational data that Tatnuck already has in spread-marts. We will also gather the NYT best seller data using the NYT Application Programming Interface (API).
Data store	We will upgrade Tatnuck's existing data warehouse and optimized the data for data analysis and query speed.
Data query and analysis tools	Here we perform data retrieval, data analysis, and data mining. We will be accessing the data store for our information.
Data presentation and visualization	This is the stage where we will represent our data using visualizations in Tableau.

(Rob & Coronel, 2009, pp. 518)

Table 3.1: Stages of Business Intelligence

People

The value of the people component of BI is to improve the experience of customers, shareholders, and employees.

Our first dashboard will simplify the reordering process for gift products by recommending when gift items should be reordered. Having accurate reorder predictions in the gift store means the gift store will be fully stocked always which will result in higher customer satisfaction, make employees lives easier, and will increase sales for shareholders.

Our second dashboard will display the Tatnuck cafe sales data, which will allow the store employees to monitor which food items they are wasting the most and which food items they are selling the fastest. This will result in a better experience for the employees and customers and it will make the employees jobs simpler and Tatnuck is more likely to have the food that the customers want in stock.

Our third dashboard will have the New York Times (NYT) best seller, and it will allow the store to monitor and order the best-selling books that customers demand which will keep the customers happy and coming back and increase sales which will make ownership happy too.

Our cafe and gift sales dashboard will use predictive analytics to provide the backroom operations staff of Tatnuck an estimate for how much gift and cafe product they should order and at what time. This system will help Tatnuck eliminate being out of stock of any gift or food item as much as possible. This system will not eliminate out of stock issues that are caused by the supplier.

Processes

Our dashboard will help Tatnuck improve their processes and become more efficient in their daily operations. Tatnuck currently has a process of removing books that have been on the shelf for more than 6 months. However, this process is very manual right now. The way they identify that a book has been on the shelf for more than 6 months in by placing a sticker on the back of the book with the date when it is first added to the shelf. Every day, a Tatnuck employee goes around and organizes the books and checks the sticker on the back of each book.

As you can tell, this is a very labor-intensive process and takes many hours each day. If Tatnuck implements a BI solution to manage their inventory, then the store employees can be notified which books need to be taken off the shelf that day and where they are in the store.

Tatnuck has a high throw-away ratio for certain products like wraps and sandwiches which we will describe in greater depth in the next section. However, by plotting the item that were being thrown away on a daily basis in a scatter plot, we were able to easily notice that sandwiches were being thrown away at a higher rate on days when less than ten items were sold. The trend becomes obvious when you throw your data into a visualization, but it would have been much more difficult to realize in a spreadsheet.

Management

The overview that Tatnuck management has of the daily operations of the bookstore will be greatly improved with our first dashboard. Our first dashboard will allow Tatnuck management to gather insights into the best-selling gift items and categories. They will also be able to predict future sales using the gift shop dashboard. As the famous Mark Twain saying goes, "The best predictor of future behavior is past behavior". By identify which items are profitable and which items are not profitable, Tatnuck can stop reordering the items that don't sell well and replace them with different items. They will also be able to easily identify the best-selling items, and then can then restock the store with items more like the best-selling ones.

In addition, having a dashboard that shows the top selling New York Times books will make managements life much easier since they will have a quick reference to the books that they should be ordering for their store. This will also allow the bookstore management to verify that the right books (NY Times best sellers) are being ordered and put on the shelf by the backroom staff.

The cafe dashboard will improve operations for management and since it will allow them to reduce waste which will decrease expenses. The biggest killers for the bottom-line profit of any restaurant / cafe is wasted food, since it's both an expense a lost revenue opportunity.

Governance

The gift items BI dashboard will assist with governance as it will allow the stores management to identify missing / stolen items. Shrinkage is a huge cost for retail stores and being able to track your inventory efficiently and identify items are that are being stolen is crucial. By knowing which items are targets of thieves, you can relocate the placing of those items so that they are closer to employees. This will help with Tatnuck's goal of decreasing expenses and increasing revenues.

Tatnuck will also gain governance by tracking gift sales, which is a key performance indicator for how well the bookstore is doing financially, through the dashboard. Having a reputable and reliable ledger for tracking sales is important especially during tax season. Tatnuck needs to verify that they are collecting sales tax and correctly reporting it.

The inventory management dashboard will also help Tatnuck with revenue recognition in their accounting process. This is crucial if they want to get a line of credit from a bank and they need to verify the amount of sales and returns they reported on their profit and loss statement. This is also critical in they get audited by the Internal Revenue Service and they to prove the amount of sales and returns they had. Tableau makes it very easy to export their sales information to a spreadsheet.

The cafe and gift sales dashboards will act as operational dashboards since they will provide an overview into the day-to-day operations of Tatnuck and allow them to track and predict sales. The benefits of the cafe and gift sales dashboards have been described in-depth previously above.

Dashboards

The NYT best seller dashboard will act as a tactical dashboard since it will allow Tatnuck to strategically order books from less well-known authors who are on the NYT best seller list and who are publishing a new book in the next six months. Many of the NYT best-selling authors have cult like followings and a lot of the less well-known authors have difficulty getting their books available in the independent bookstores. By implementing our NYT dashboard, Tatnuck can become the one stop bookstore to purchase books of less well-known authors, they can cultivate the relationship among avid book readers in the central Massachusetts area. Also, avid book readers tend to be excellent repeat customers, and they also tend to spend a lot of money on coffee and snacks, which will increase cafe spending. The NYT dashboard has a two-pronged effect by increasing new book sales which will also lead to increased cafe sales.

To take the NYT best sellers' dashboard one step further, Tatnuck can devote an entire section of the store to books on the NYT best sellers. People often tend to gravitate to books written by NYT best-selling authors since there is a level of notoriety that comes along having your book on the

NYT best sellers list. Housewives also tend to enjoy NYT best sellers' books since they are great conversation starters.

The cafe and gift store dashboards are both operational and tactical since they will allow Tatnuck to have a better understanding of what is going on in their business from a day-to-day perspective and allow them to make adjustments to their product offerings to improve sales.

Data Warehouse Structure

We are assuming that Tatnuck already has an existing data warehouse and that it just needs some upgrades. It is crucial that every business develops a Star Schema plan before actually building out their data warehouse. Therefore, we have created the Star Schema diagrams for both the cafe and gift sales to show how we would structure the data in the updated data warehouse for Tatnuck.

For the cafe and gift sales dashboards, we will use Tatnuck's internal cafe sales data and their internal gift sales data. The only dashboard that we will use external data is the NYT best seller dashboard. The NYT best sellers are extremely easy to download using an Application Programming Interface (API) that the NYT has made publicly available.
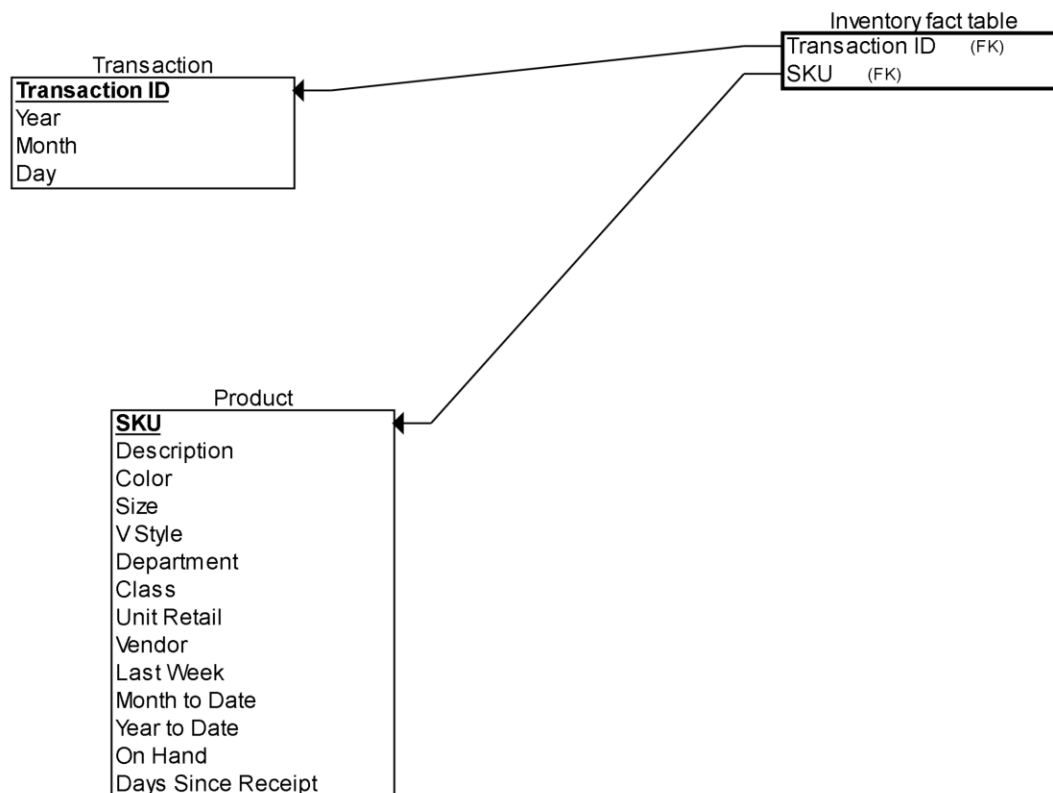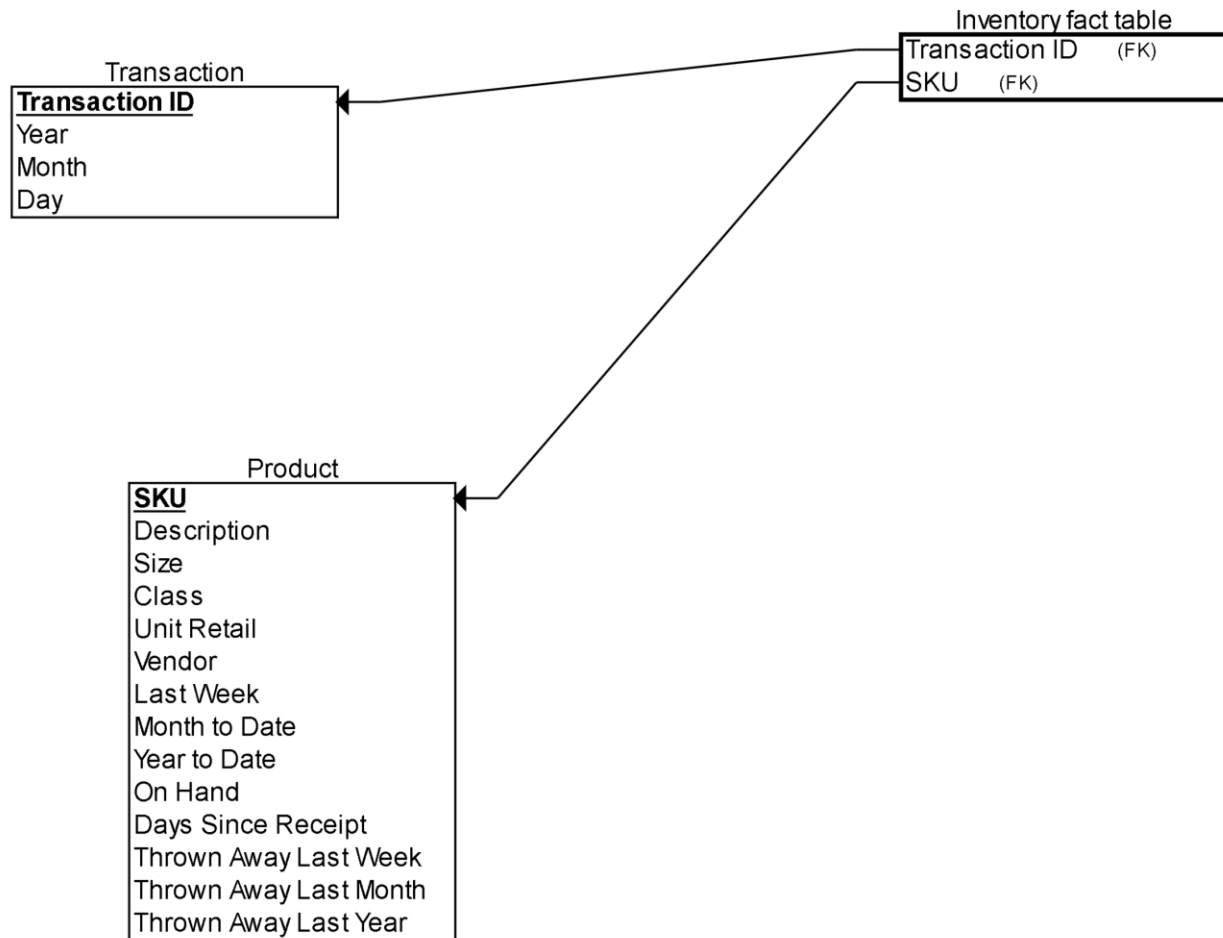
Figure 4: Star Schema for Tatnuck Cafe

Figure 5: Star Schema for Inventory

Prototypes

Currently the computerized/automated tools which Tatnuck Booksellers is using are not intuitive and lack visualizations. Therefore, we are proposing a BI solution that will present their data in more meaningful ways. Tableau is one interactive data visualization BI tool that can transform the data into live dashboards and reports. Below we present three prototypes: two dashboards developed in Tableau and data analytics.

Dashboard 1 - Cafe and Gift store Department Dashboard

This mix of operational and tactical dashboard was built for the gift store and cafe managers. It focuses on six types of information from business results: items wasted, item sold, sales of products for days of the week, sales of the products with maximum temperature of the days of week, the quantity of products been sold for each week and quantity of the products sold from each departments.
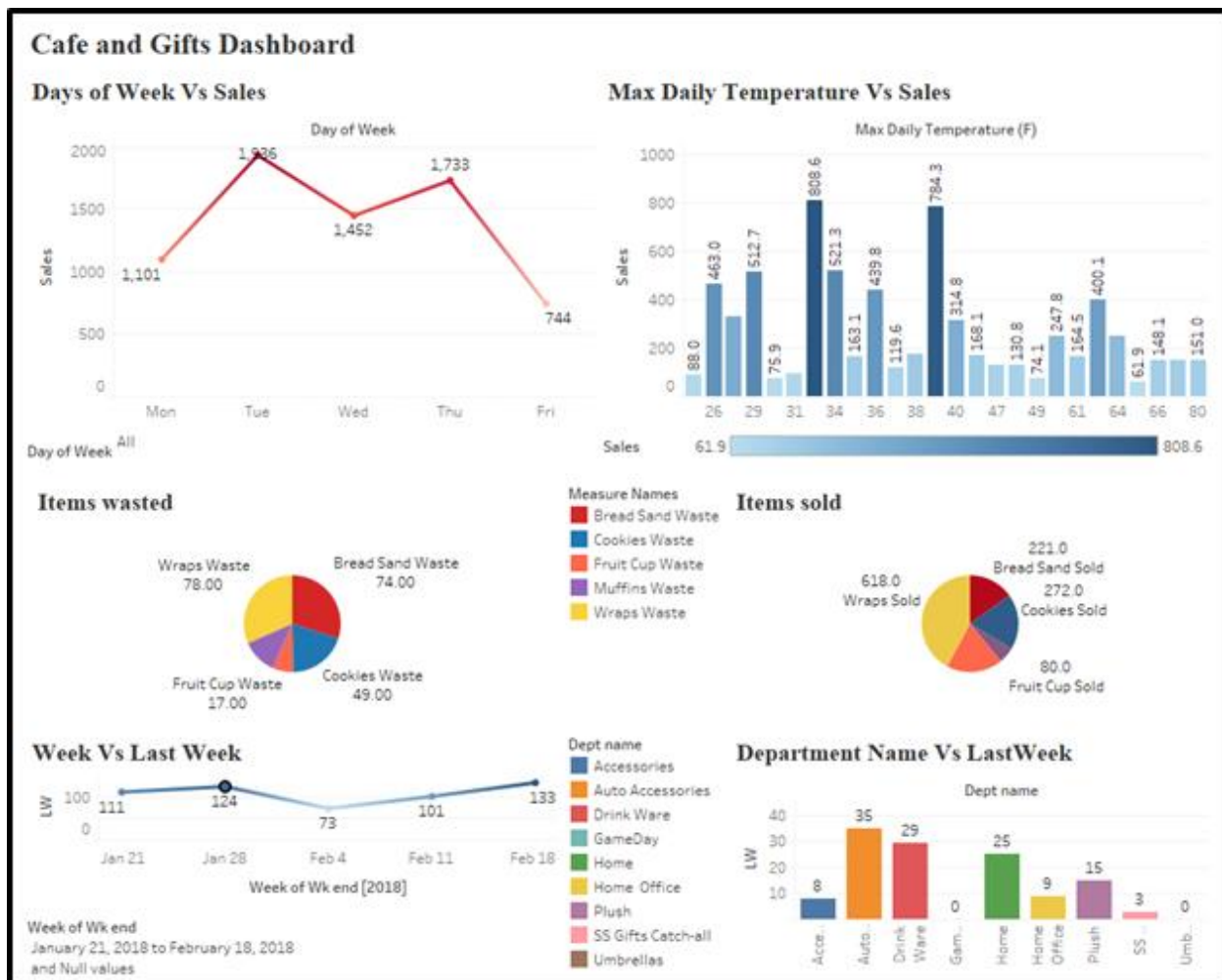

Figure 6: Cafe and Gift Store Dashboard

The dashboard will inform the manager insights regarding three main aspects.

1. For cafe, the dashboard can tell what the maximum sales of the products is for days of week.
2. For café products, the dashboard can tell which products the mostly sold and which items are being wasted.

3. For the products, the dashboard can also tell the sales of the products according to the temperature for days of week.
4. For the gift items, the dashboard can tell the quantity of products been sold during each week and from which department.

Days of week Vs Sales

The sales of the products is used as KPI to show the sales of products for days of week. The user can filter the day to see the sales of the product for a specific day. The user can also click the sales of the product to see the items sold and items wasted. By presenting the items been sold and wasted, the manager would have a knowledge about what items are sold, items wasted and sales of the products.
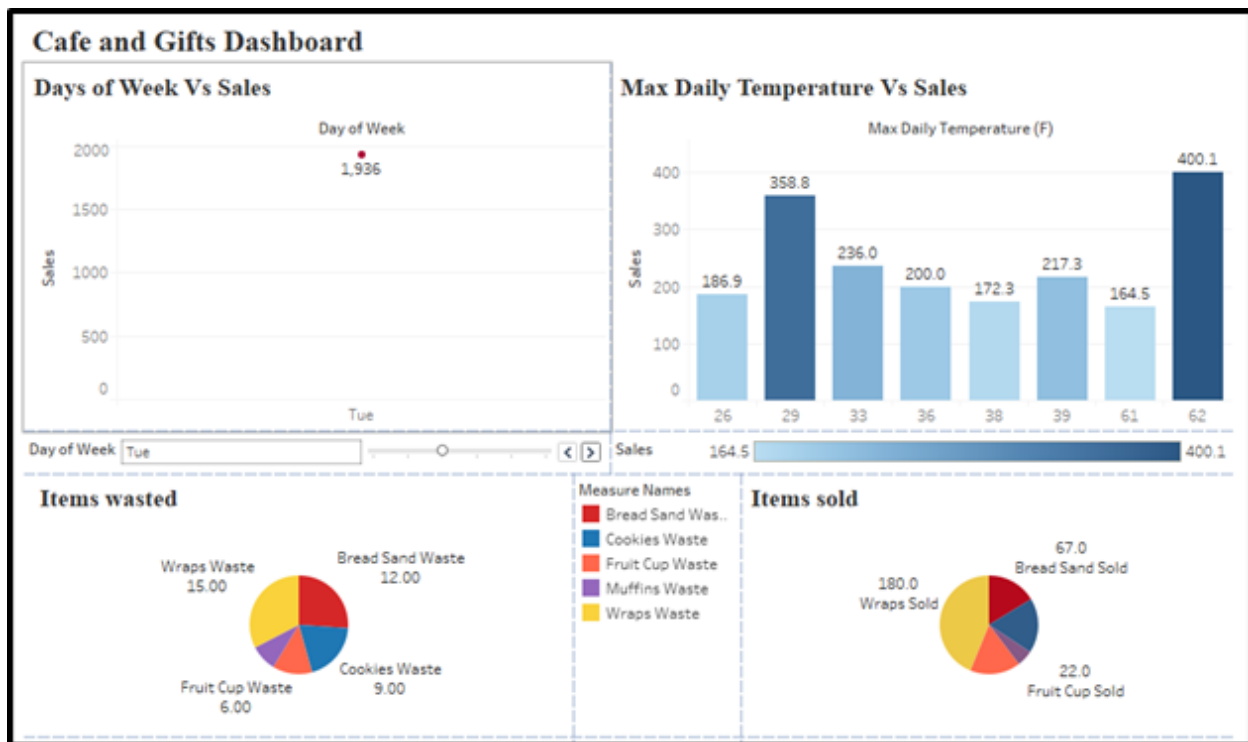

Figure 7: Cafe and Gift Store Dashboard - Cafe Section

Maximum Daily Temperature with Sales

The dashboard shows the sales of the products during the temperature. It is an overview on how the sales of the products are affected during different temperatures. By clicking on specific temperature, the information in other charts changes accordingly.
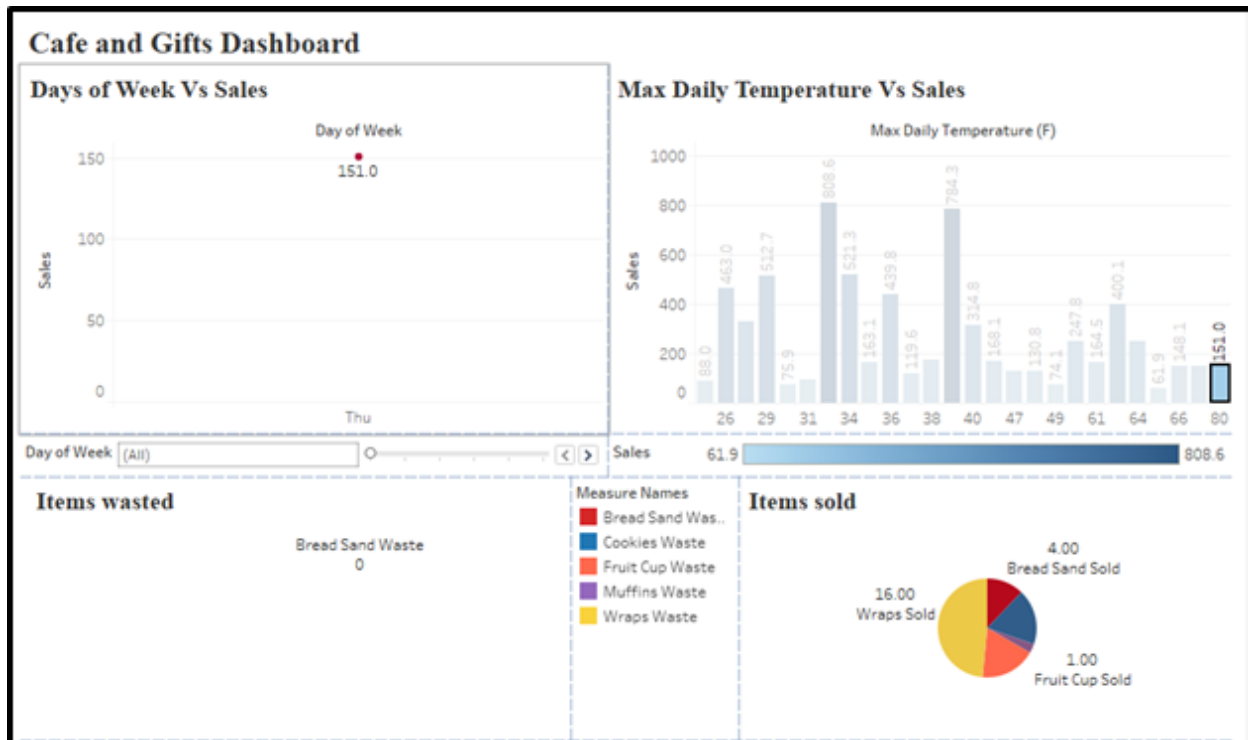
Figure 8: Cafe and Gift Store Dashboard - Cafe Section

Quantity of the Products Sold

The dashboard provides the information about the quantity of the products that are sold during each week. We have the data from January 21- February 28. The chart allows the user to select and filter the information based on a specific week. If the manager wants to see the quantity of the products sold during each week, he can just select a specific week. The manager can also see the trends of sales during specific week.
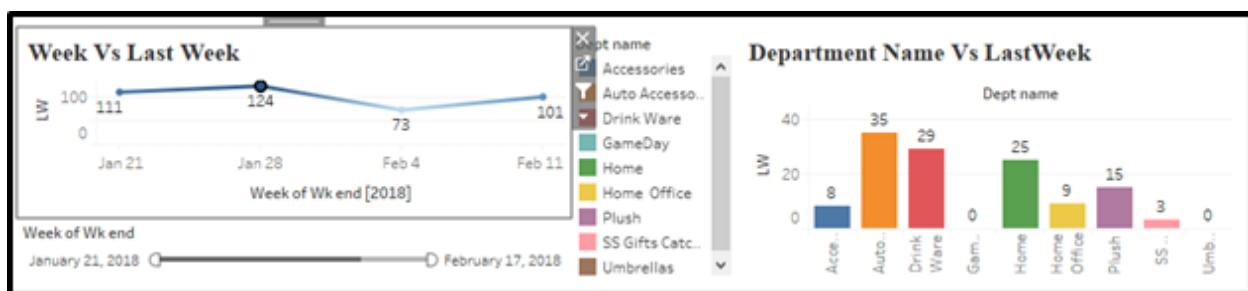

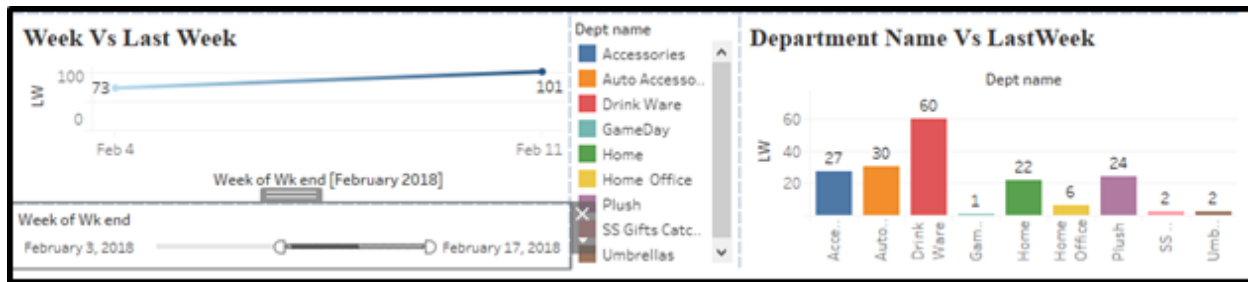*Figure 9: Cafe and Gift Store Dashboard - Gift Store Section*

Figure 10: Cafe and Gift Store Dashboard - Gift Store Section

Products sold from each department

The dashboard presents the products sold from each department during each week. It provides the information about the products sold from each department so that manager can restock the items. By clicking the department name, the related information about that department will reflect in the rest of the charts. The products from the departments include ceramic mugs, travel mugs, water bottles and so on.
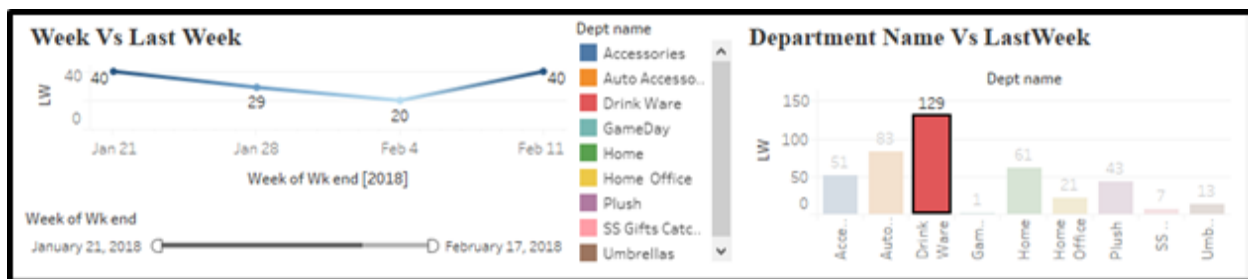
Figure 11: Cafe and Gift Store Dashboard - Gift Store Section

Dashboard 2 - New York Times Best Sellers Dashboard

The New York Times Best Sellers (NYTBS) dashboard is intended for use by the bookstore manager. In short, it allows the user to execute various queries related to books on the NYTBS list and foresee authors on the list who have new books coming out in the near future. Before going into the operational details of the dashboard some preliminaries will be presented.

The first step in building the NYTBS dashboard was to acquire the appropriate datasets. In total, two datasets were used: the NYTBS dataset and a Forthcoming Author dataset. To facilitate the use of their data, the New York Times has created an API (New York Times Developer Network, n.d.) that allows the downloading of NYTBS list data. In order to interface

with the API, Python code was written to query the NYTBS list data and write it to a csv file (see appendix 8.1). Once the data was in csv format it could be utilized in a dashboard.

The Forthcoming Author dataset was not a real dataset and its contribution to the dashboard was simulated. However, a dataset containing information on forthcoming books and their associated authors is surely available to a bookseller such as Tatnuck. Thus, as a proof-of-concept the simulated dataset produced the desired effect. To model the integration of the two datasets 10% of the authors in the NYTBS dataset were randomly selected to have a book coming out in the next six months. This six-month window was selected for demonstration purposes and is an adjustable parameter in the eventual fully operational dashboard.

The NYTBS dashboard, shown in figure 12, consists of four quadrants in the main part of the dashboard and three filters in the top right. The Book List (top left) displays textual information about the books on the NYTBS list, separated by category, weeks on list (WOL), their rank in the respective category and title. The author bar chart (bottom left) ranks authors by the sum total of weeks all their books have been on the NYTBS list. Similarly, the titles bar chart (bottom right) ranks titles by the same sum total of weeks the title has been on the NYTBS list. Lastly, the categories circle chart (top right) displays which categories have the most total weeks on the NYTBS list for all book in each category; with the size of the circle proportional to the sum total of weeks on the NYTBS list. Note that placing the mouse over any of the circles will display the category name. Thus, although the smaller circles do not have their names printed this information is easily accessible. The three filters in the top right corner allow the user to filter the category, rank and select whether the authors, titles, etc. appearing on the dashboard has a new book coming in the next six months. The next three sections demonstrate the interactive features and present three targeted uses of the NYTBS dashboard.
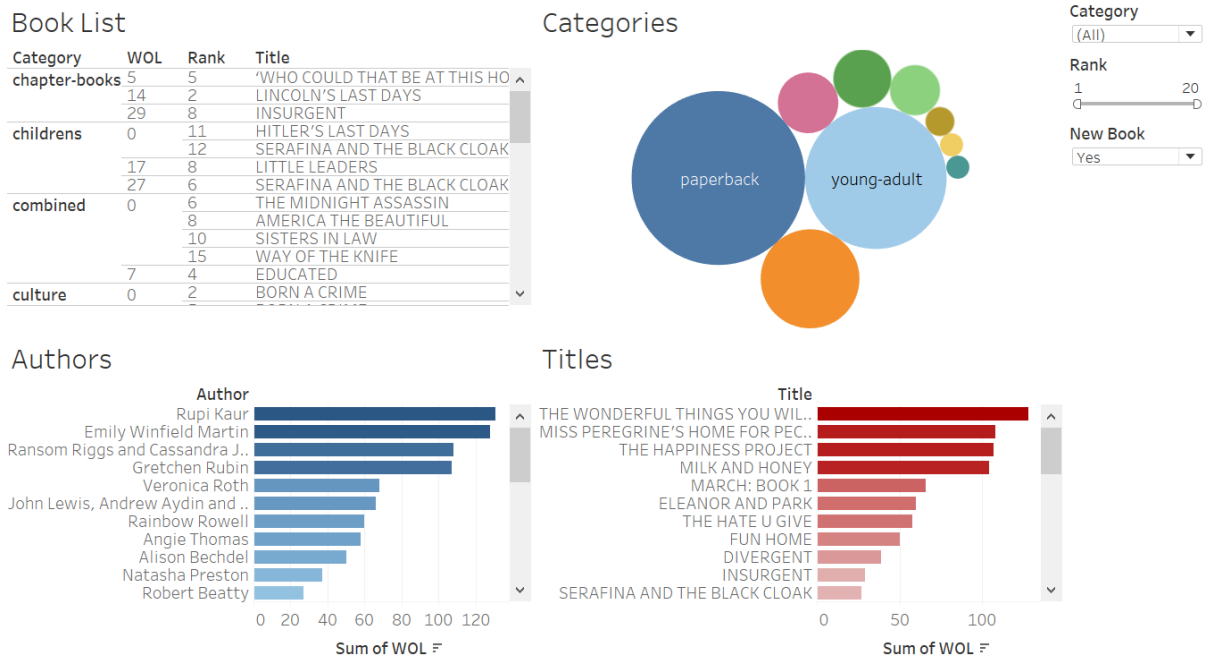
Book List

Category	WOL	Rank	Title
chapter-books	5	5	'WHO COULD THAT BE AT THIS HO
	14	2	LINCOLN'S LAST DAYS
	29	8	INSURGENT
childrens	0	11	HITLER'S LAST DAYS
		12	SERAFINA AND THE BLACK CLOAK
	17	8	LITTLE LEADERS
	27	6	SERAFINA AND THE BLACK CLOAK
combined	0	6	THE MIDNIGHT ASSASSIN
		8	AMERICA THE BEAUTIFUL
		10	SISTERS IN LAW
		15	WAY OF THE KNIFE
	7	4	EDUCATED
culture	0	2	BORN A CRIME

Categories

Authors

Titles

Figure 12: New York Times Best Sellers Dashboard

Filtering by Category

All four quadrants and three filters are interactive and interconnected. For example, selecting 'category' in the filter will display information in each quadrant related to books only in the business category, as shown in figure 13. This feature may be useful for the book manager who wants to create a themed display focusing on a particular category or topic, such as business or nature. This could facilitate in-store NYTBS focused displays on a specific category or a variety of related categories.
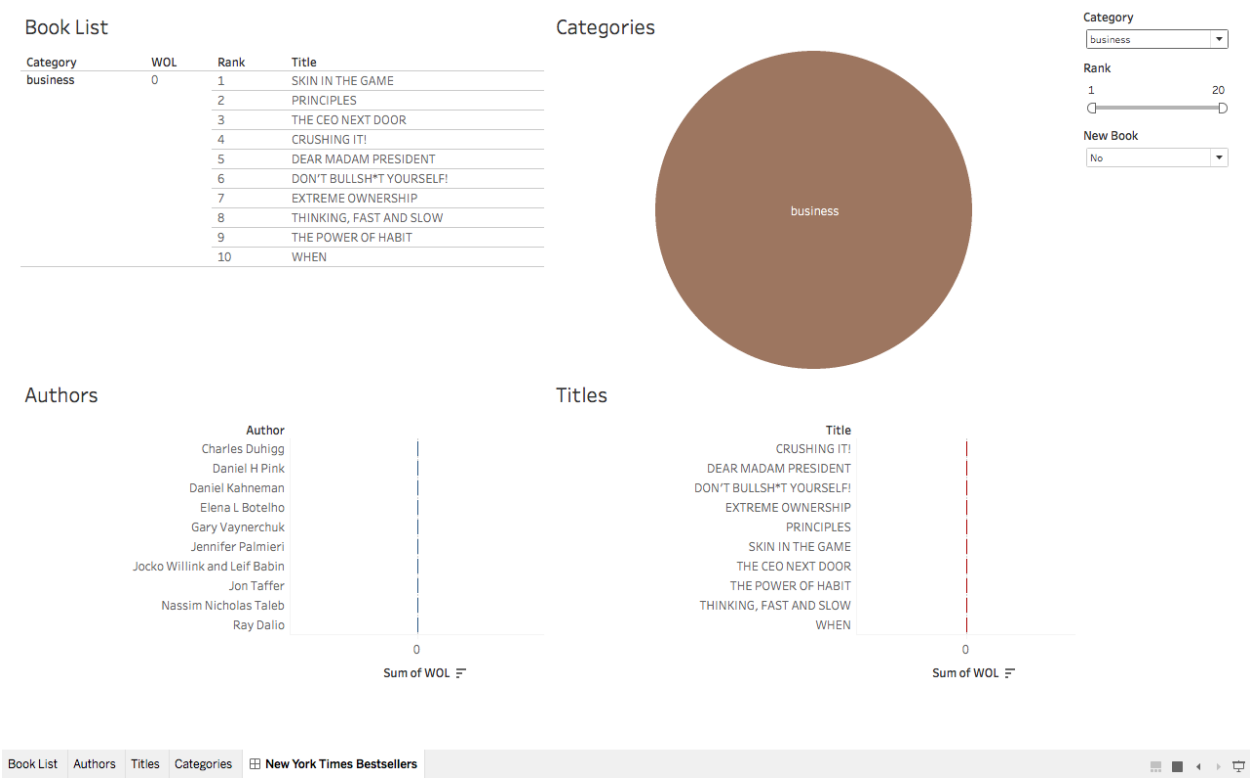
Book List

Category	WOL	Rank	Title
business	0	1	SKIN IN THE GAME
		2	PRINCIPLES
		3	THE CEO NEXT DOOR
		4	CRUSHING IT!
		5	DEAR MADAM PRESIDENT
		6	DON'T BULLSH*T YOURSELF!
		7	EXTREME OWNERSHIP
		8	THINKING, FAST AND SLOW
		9	THE POWER OF HABIT
		10	WHEN

Categories

business

Category
business

Rank
1   20

New Book
No

Authors

Author	
Charles Duhigg	
Daniel H Pink	
Daniel Kahneman	
Elena L Botelho	
Gary Vaynerchuk	
Jennifer Palmieri	
Jocko Willink and Leif Babin	
Jon Taffer	
Nassim Nicholas Taleb	
Ray Dalio	

Sum of WOL

Titles

Title	
CRUSHING IT!	
DEAR MADAM PRESIDENT	
DON'T BULLSH*T YOURSELF!	
EXTREME OWNERSHIP	
PRINCIPLES	
SKIN IN THE GAME	
THE CEO NEXT DOOR	
THE POWER OF HABIT	
THINKING, FAST AND SLOW	
WHEN	

Sum of WOL

Book List   Authors   Titles   Categories   ⊞ New York Times Bestsellers

Figure 13: New York Times Best Sellers Dashboard, Filter By Category

Filtering by Rank

Using the rank filter, the user is able to customize the range of ranks on the NYTBS list as shown in figure 14. This could be used to create a NYTBS themed display with the top authors for each week. As there are certainly readers who select books based on the NYTBS list, a display focusing on top ranked books is another possible NYTBS list themed in-store display.
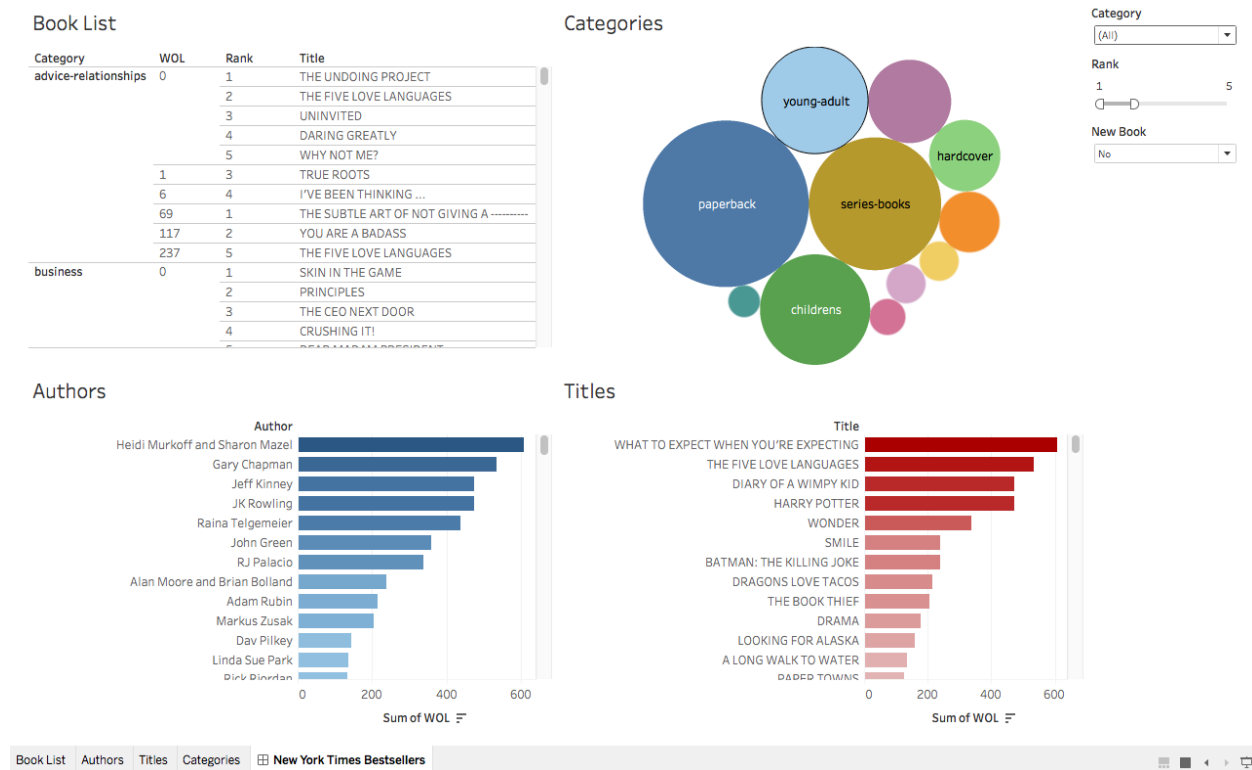
Figure 14: New York Times Best Sellers Dashboard, Filtering By Rank

Filtering by Forthcoming Book

The New Book filter allows the user to display results in the four quadrants based on whether an author has a forthcoming book in the next six months. This filter has three options: 'All', 'Yes' and 'No'. Selecting 'All' displays all authors, titles, etc. in the four quadrants regardless of whether the corresponding author has a forthcoming book. 'Yes' displays only authors, titles, etc. which have a book forthcoming and 'No' authors, titles, etc. which do not have a forthcoming book. The primary purpose of this feature is to allow the user, i.e. book manager, to determine authors who are on the NYTBS list and have a new book coming out. However, this feature would be optimal if the user were able to determine authors new to the NYTBS list (that is with a low number of weeks on the list) which have a forthcoming book. The reason being authors that have been on the NYTBS list most likely are well known to book store managers. By selecting 'Yes' in the new book filter and scrolling to the bottom of the authors bar chart, the user can find authors newly added to the NYTBS list with a forthcoming book. Additionally, when the user clicks on the specific title in the book list a pop-up window with any hyperlinks corresponding to the particular title appear. Some examples of possible hyperlinks in the NYTBS dataset are amazon page, Sunday New York Times book review and sample chapters. All these above features are displayed in figure 15.
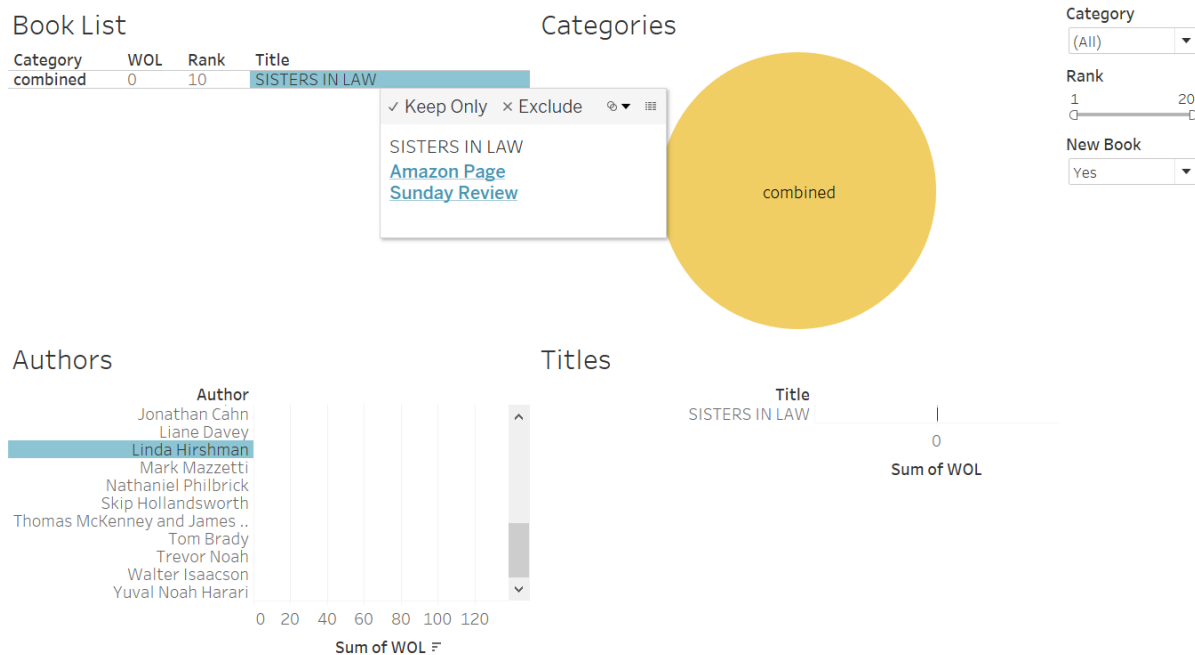
Figure 15: New York Times Best Sellers Dashboard, Author with Forthcoming Book

Data Analytics

The dashboards provide dynamic visualizations and interactivity as to how the company was performing in the past with sales, inventory and customers data and how it has been in terms of current transactions. However, it orders to understanding trends in the past performances and to generate recommendations from the current data, we perform predictive analytics on the datasets. Our main aim concentrating on analytics is to find peculiar trends, recommend plans for the future and generate customized marketing.

Tatnuck is divided into 3 division of provisions which are bookstore, cafe and gift store. The cafe and gift store are introduced to provide better experience to customers and host several events and help the store go beyond the conventional bookstore definitions. Gift store and cafe allow us to understand several trends in the performance as they are directly related to the customers and their presence and buying behavior in tatnuck. Therefore, we have performed analytics on these divisions to understand buying habits on a daily basis and preferences of the customers.

The cafe data contains 60 observations as part of 2 months of data in terms of sales for each product, items wasted, temperature and the corresponding day. The dataset required minimal cleaning since it consisted of aggregate data from each day. The initial step of analytics was to understand the importance of each variable. In order to do this, we first needed to remove some

variable that could bias the results and as a result we removed Serial Number, Date, Day Code and Day Name. Removing these variables, the resultant dataset contained only factors which affect the sales in terms of products and temperature. Thus, we understand the effect each of the factors on the other.
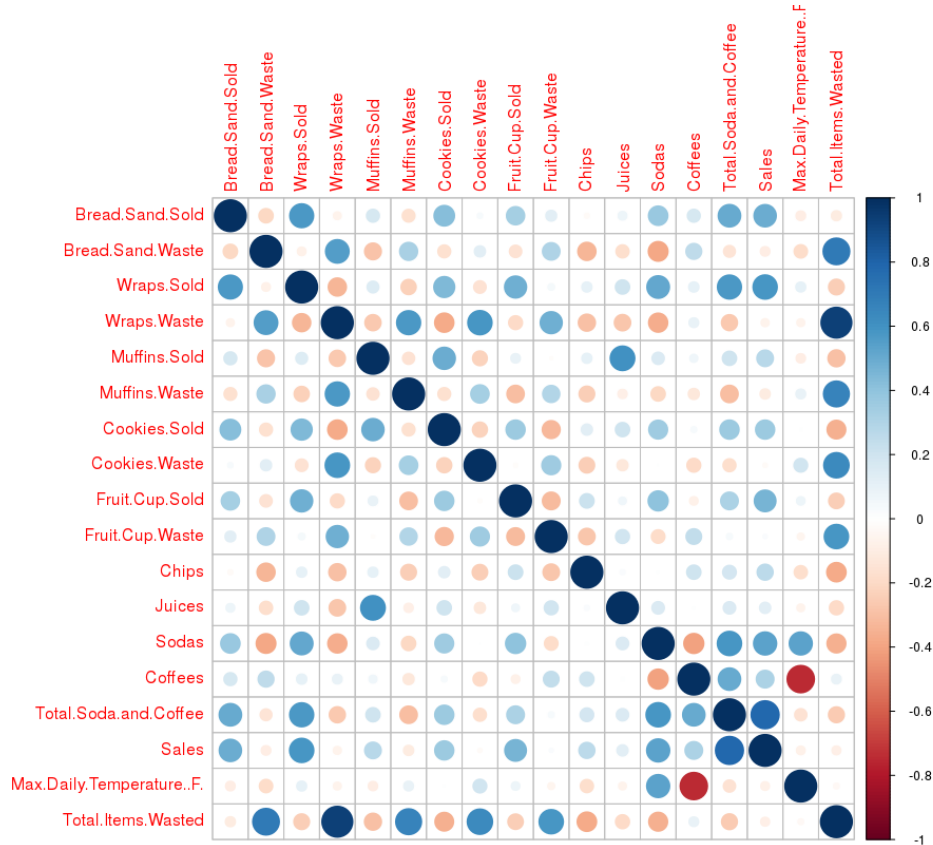

Figure 16: Correlation Matrix

The graph dynamically displays the dependency of one factor over the other. The darkness of the dots displays the degree of dependency of one factor with the other. The darker the dot is the higher the one thing depends on other and this is called as correlation. For example, we can see that Total Items Wasted and Total Items Wasted have extremely dark circle. This is because they are understandably the same things. Thus, they must directly depend on each other, in other words they must have perfect correlation. A blue dot means that as one thing increases, the other also increases and a red dot means as one thing increases the other decreases.

We see several trends from the graph above but particularly interesting and worth understanding trends are the relation between total items wasted and wraps wasted and the relationship between temperature and the coffees sold. The first shows that total items wasted depends very strongly on the wraps wasted. Therefore, the wraps are a major contributor to waste and we would examine

this in detail. The other relationship is between temperature and coffees sold. As the temperature increases the coffees sold decreases. We need to understand this since coffees are one of the highest contributors to sales. Therefore, in this section we would concentrate on these two trends in detail.

The blue dot in the area connecting total wasted items and coffees shows that as the wraps wasted are a major contributor to the total wasted items. This could mean several things. It could mean that the supply of wraps is greater than the demand. Also, it could be possible that the wraps are not a popular product and cause severe losses to the cafe. In order to understand this, we look at the how the wasted products differ over the total wasted items.
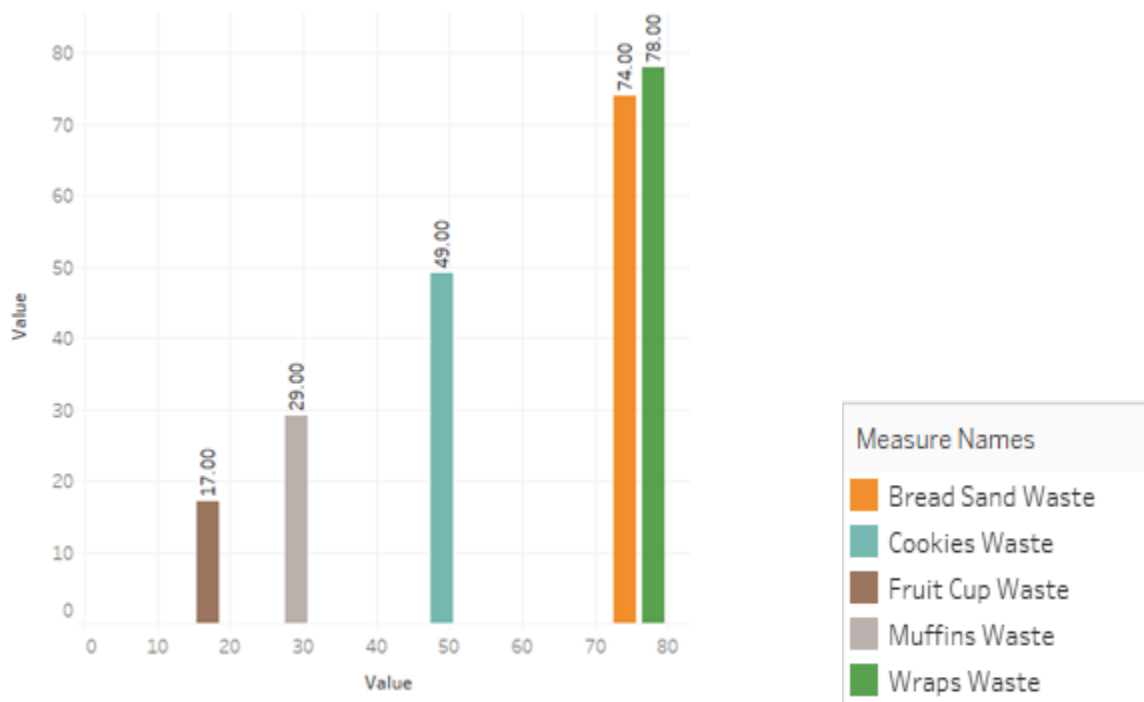

Figure 17: Products Wasted

The graph above shows the total products wasted and differentiates this based on the products. The graph shows that wraps are the highest contributor to wasted items with 78 wraps wasted across the two months. However, bread sandwiches are a closed second. The total wasted sandwiches are 74 across 2 months. This means that we need to look at both the items together since both of them are significantly contributing to waste. Thus, we start by looking at the wraps and their effect on the total items wasted.
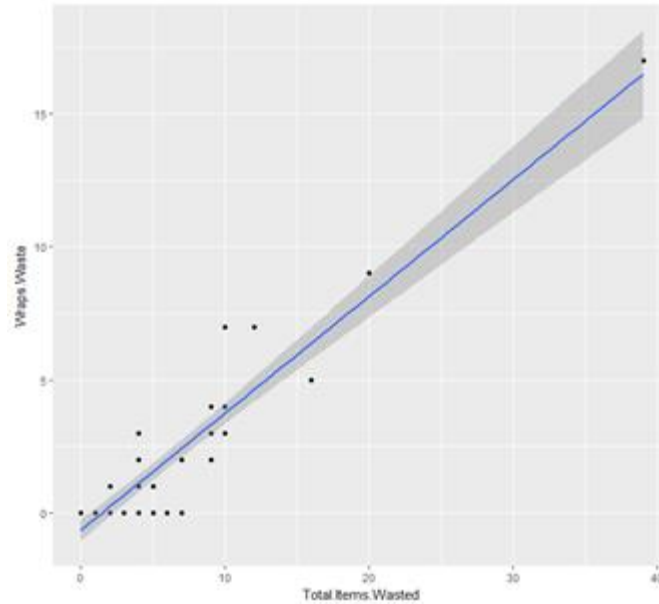
Figure 18: Total Items Wasted vs Wraps Wasted

The plot agrees with our expected results in the fact that wraps are a huge contributor to the wasted items. But more importantly, we can view the power of analytics through this plot. The point on the rightmost corner of the plot is called as an outlier in the term of analytics. Outlier is a point which is extremely rare but can cause severe effects on the data. It is very important to understand this point to avoid this trend in the future. It may be possible that the point displays a holiday on which the stocks were kept high and resulted in losses and other factors like that. Thus, it is extremely important to remove these points.

Now, since bread sandwiches were a close second it is important to understand the effect of each of these products in terms of the number of products sold to the number of items wasted. In order to understand this, we plot the following graph which shows the products in terms of items sold to items wasted.
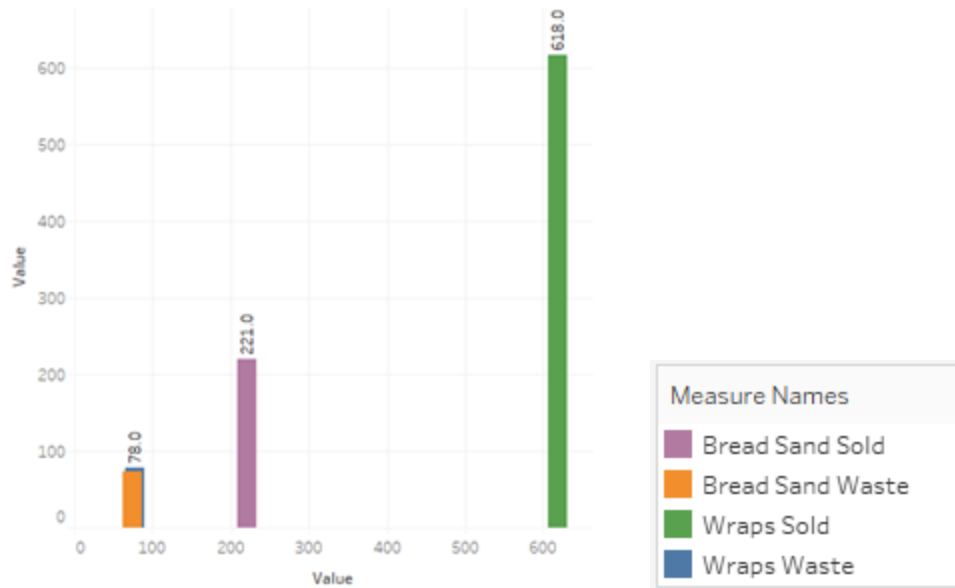
Figure 19: Bread Sandwiches and Wraps Sold and Wasted

The plot above shows another unique trend. The initial hypothesis which said that wraps were the most wasted items was indeed true however looking at the above graph, the wraps are extremely famous, and the number of wraps sold are extremely high making it the most popular item on the menu. On the other hand, the bread sandwiches are extremely low selling item. The graph shows that out of every 3 bread sandwiches made, one is wasted. In this case, the sandwiches are in a critical situation and therefore the cafe must take adequate steps to overcome this wastage of sandwiches.

In order to understand this extreme wastage of sandwiches, we look at how the total wasted items depends on the sandwiches.
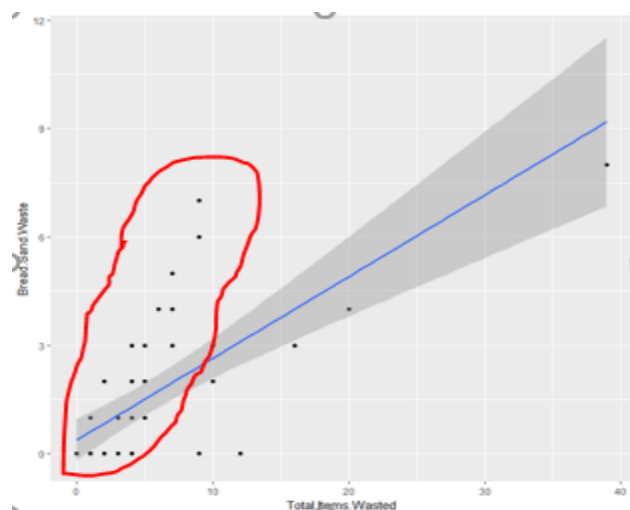
Figure 20: Hidden Trends in Bread Sandwiches Wasted

The red circle indicates an alarming trend in the sandwich sale. During these days, where the maximum items sold are less than ten, the total sandwiches wasted are close to 8. This means that out of 10 products wasted 8 were sandwiches which makes sandwiches extremely unpopular. It is very important to look at this trend and understand it.

Looking at all these trends and hidden information obtained from the dataset, we provide three robust recommendations to the client. First, it is highly possible that the cost of bread sandwiches is extremely high. The dataset provided gives the sales in numbers of items not amount in dollars. Thus, we assume that the cost is very high and recommend it to bring it down to an optimum value. Second, package wraps and sandwiches with other product to increase sales and reduce the costs of wastage. Provide offers with new customized versions and customer specific meals. Third, provide timely offers on the lesser selling items as it is important to get the sales up from wasted rather than increasing new sales.

Now, let us look at the correlation matrix again.
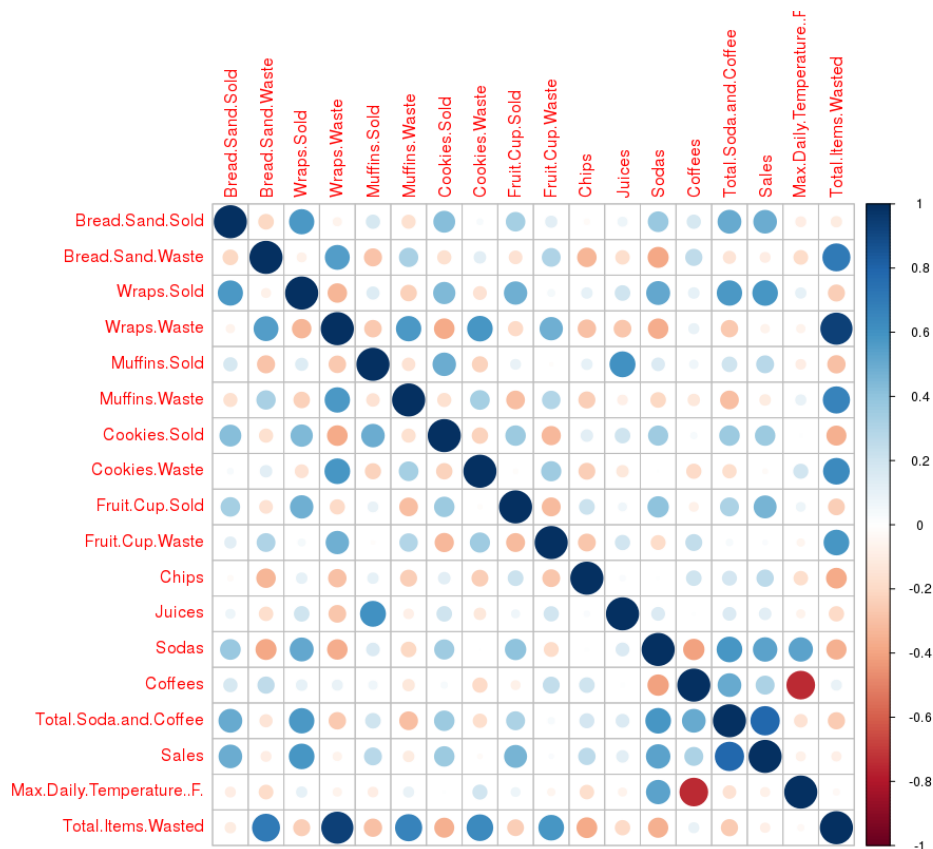
Figure 21: Correlation Matrix for finding Trends in Temperature

We now want to analyze the effect of temperature on the number of coffees sold. The coffees sold in the cafe as per the correlation matrix decreases as the temperature increases. We want to look into this in detail since coffee is one of the largest sellers in the cafe.
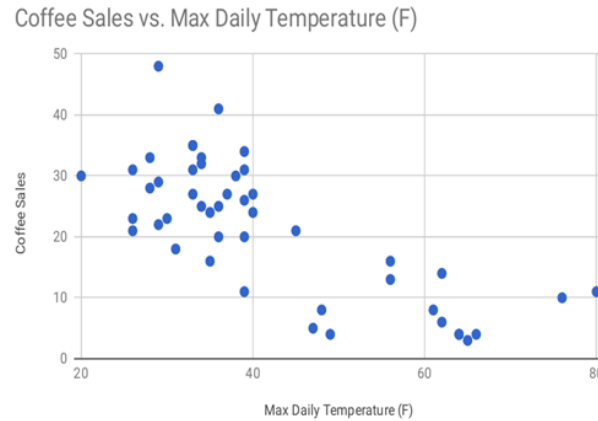

Figure 22: Temperature vs Coffee Sold

The figure shows that the number of coffees sold indeed depends on the temperature. In fact, the total number of coffees decrease by 1 on average with every 3 degree increase in maximum temperature. This is highly undesirable since coffee is one of the largest sellers and the largest companion to book reading. Thus, due to this, we have come up with 3 recommendations that could help boost up the sales with the temperature increase and also understand sales of coffee in detail.

First, the cafe only sells hot coffee. It is thus understandable that as the temperature increases the demand for hot coffee decreases. Therefore, we see a significant negative trend in this plot. In order to bring sales to an optimum level even in summer, we recommend the client to offer iced coffees. Along with offering iced coffee maintain the inventory in such a way that there is minimal wastage of both types of coffees with different temperatures. Second, gather point of sales data. We want to understand the time of day at which the customers are in huge numbers in the store. Provide offers at this time, generate a great atmosphere for reading books in the store so people could take coffee with it. Third, monitor inventory with the change in temperature. The store would not want wastage of hot coffee at high temperature and cold coffee at low temperature and therefore we recommend the client to understand the compromise of inventory levels to optimum level.
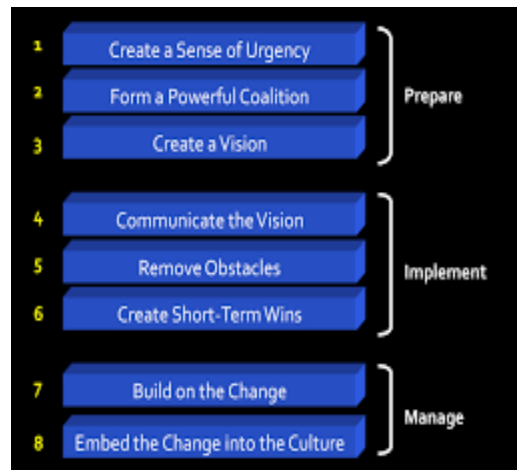
Implementation

Figure 23: Kotter's 8 Step model

Managerial Implementation

There are several challenges that Tatnuck would need to overcome for implementing a BI solution. We will use the 8-step model given by Dr. John Kotter, as a reference for the managerial aspects of implementation of a BI solution.

Step 1: Create a sense of urgency:

Tatnuck may face lack of visibility if they do not have a BI system and may not be aware of potential threats in the future. It is necessary to highlight the underlying crisis in terms of unhappy employees, rising costs and having competitive edge to keep the balls rolling. Tatnuck doesn't have clear visibility for tracking of books, sales of gifts and café, so it is essential to create sense of urgency amongst the stakeholders. Urgency in terms of Tatnuck is to have a competitive edge over Barnes and Nobles and Amazon.

Step 2: Form a powerful coalition:

Once the sense of urgency has been created, it is important to put together a powerful team, to drive this change in the organization. The team should have a group of skilled people, significant level of trust amongst themselves and shared objectives.

Step 3: Create a Powerful vision:

Here it is important to demonstrate how the future will be different from the past. Some of the important characteristics that the vision should have are that, it should be imaginable, desirable, feasible, focused, flexible, and communicable.

Step 4: Communicate the vision:

Once the vision has been created, it should be applied to all operations in the company. It should be discussed openly amongst all the employees. It is important that the communication of this vision should be simple, vivid, repeatable and invitational. It should be very simple without the use of high-level technical jargon that people would find difficult to understand. The company should use verbal pictures, metaphors and analogies to explain. The ideas should be communicated to everyone in the organization.

Step 5: Remove Barriers:

It is important to remove any obstacles to change, change the system that would potentially challenge the vision and encourage risk taking. It is important to show why this change is needed, which would build a sense of positivity and self confidence amongst the employees.

Step 6: Create short term wins:

Visible and definite steps must be created within the organization very early on during the project. This would help to validate the vision and strategies. It is important to reward individuals during this stage, who can meet their targets. This would give the people working hard, recognition and encouragement.

Step 7: Build on the change:

During this process, it is important to maintain this momentum and do not give up after achieving short term success. In the process, critical momentum may be lost and potentially followed up with attacks by individuals within the organization that still resist the change initiative. This can be done by nurturing the positive values and keeping the ideas fresh. The company should follow the process of continuous improvement.

Step 8: Embed the change into the culture:

The change initiatives must be embedded into the culture of the organization for it to be a long-term success. Continuous efforts should be put in to make sure that the change in seen in every aspect of the organization. The company leaders must support the change by reinforcing the new culture through training and coaching, telling vivid stories about the organization, what it does and why it is successful.

Technical Implementation

Implementing a BI system will improve business outcomes. It is a principal applied to technologies, helping organizations make sense of data. Implementing a BI system will improve business outcomes. It is a principal applied to technologies, helping organizations make sense of data.

Currently Tatnuck Bookseller has an imerchant software that handles the company's data and uses excel to generate Ad-hoc reports.

With our implementation of Business Intelligence solution, we created Dashboard on Tableau to extract data and presents it to the manager to make meaningful conclusions. Based on this implementation of our Business Intelligence solution, we have highlighted technical challenges for Tatnuck Booksellers as below:

IT infrastructure

Tatnuck Bookseller doesn't have IT infrastructure, so IT infrastructure must be set up for the BI solution to be deployed. As our BI solution would require processing and analyzing large amount of data, it would need an efficient IT infrastructure in place that can accelerate analysis. If the hardware, software and network resources for installing, deploying and running the BI solution is not best the users will struggle to understand the system and find it harder to enhance the skills for using BI solution.

Implementing an efficient IT infrastructure will help the managers to identify the business problems at ease, scalability of business in future. The good IT infrastructure would be really very helpful for expansion of business in future.

Data Integration

BI Solutions usually require huge amount of data to analyze and produce results and data management has always been a tricky job for IT to handle. With the growth of database in an organization moving data between applications and departments increases complexity. Thus, with this move Tatnuck needs to make sure that they have understood their data needs for implementing the BI solution. The right data needs to be reported or delivered to the right department and applications.

As the organization advances there can be changes in the data structure set, thus, Tatnuck needs to consider that with every change in the data structure it needs to be associated with the integration of data with the BI solution implemented.

Data Quality

A business does not want to produce inaccurate reports, and most of the times data input to a Business Intelligence system is unstructured. Thus, Tatnuck needs to understand how vital it is to maintain and observe the data quality if they are looking for implementing the BI solution. An inaccurate data will create problems for users when running BI solution. It is an effort as it would require people to spend huge amount of time in structuring and cleaning data for efficient usage and analysis. The backend of the solution should be efficient enough to handle data before it is input to the Business Intelligence system. This is the only way to ensure that the front end of the BI solution would generate accurate graphs, reports and dashboard.

As any information that a company gathers from the data is limited to the quality of data. Thus, the data provided to the system has to be of the highest quality. This would require a lot of time investment from Tatnuck to streamline the data for the proposed BI solution.

Conclusion

1. Implementing analytics platform can reduce all the manual tasks and improve efficiency.
2. Automating this process will reduce human errors resulting in more accurate prediction.
3. Easy monitoring of real time data will help in re-stocking and reducing wastage. You can monitor your sales of the products daily and restock the items in advance and reduce the wastage.
4. Analyzing data will help in reacting quickly to changing demands of customers. For example: when the temperature is low the sales of coffee are high Tatnuck can introduce different types of coffees with discount rates or coupons.
5. A comprehensive BI system will improve visibility into the important operations in the organization. The BI software can also be used to identify the areas that need improvement. This will also save a lot of time and improve productivity of employees using this software.
6. A BI system can also be used to make successful strategic plans for the organization. It can help translate raw data into actionable information and identify any suitable opportunities or potential threats for the organization, and hence enable to plan for a successful future.

References

• Tatnuck Bookseller (n.d.) Retrieved from www.tatnuck.com

• New York Times Developer Network (n.d.) Retrieved from https://developer.nytimes.com/

•Tableau Training Videos (n.d.) Retrieved From https://www.tableau.com/learn/training

•R Training (n.d.) Retrieved From https://www.r-project.org/help.html

•Dashboard Training (n.d.) Retrived From https://onlinehelp.tableau.com/current/pro/desktop/en-us/dashboards_best_practices.html

•Awesome Tableau Dashboards (n.d.) Retrived From
https://www.pinterest.com/catmule/awesome-tableau-dashboards/

[7] Clancy, H. (2016, August 8) *This Is Analytics Software Firm Tableau's Strategy for Winning Bigger Deals* Retrieved From http://fortune.com/2016/08/08/tableau-bigger-deals/

[8] Alteryx Analytics Reviews (n.d.) Retrieved From
https://www.trustradius.com/products/alteryx/reviews/pros-and-cons?cg=large

[9] Rob, P., & Coronel, C. (2009). *Database systems: Design, implementation, and management* (pp. 517-518). Boston, MA: Course Technology.

Gaut, R. (2015, November 16). User Adoption and ROI Go Hand-in-Hand. Retrieved From
https://www.csgpro.com/blog/2015/11/user-adoption-and-roi-go-hand-in-hand130108

8. Appendices

8.1 Python Code to Interface with New York Times Best Sellers List API

```python
import numpy as np
import os
import csv
import requests
from datetime import datetime
from time import sleep

# Parameters
api_key = '*** PLACE API KEY HERE ***'
base_url = 'http://api.nytimes.com/svc/books/v3/lists/'
date = 'current/'
overview = 'overview/'
response_format = 'json'

list_names = [
                'combined-print-and-e-book-fiction',
                'combined-print-and-e-book-nonfiction',
                'hardcover-fiction',
                'hardcover-nonfiction',
                'trade-fiction-paperback',
                'mass-market-paperback',
                'paperback-nonfiction',
                'e-book-fiction',
                'e-book-nonfiction',
                'hardcover-advice',
                'paperback-advice',
                'advice-how-to-and-miscellaneous',
                'chapter-books',
                'childrens-middle-grade',
                'childrens-middle-grade-e-book',
                'childrens-middle-grade-hardcover',
                'childrens-middle-grade-paperback',
                'paperback-books',
```

```
                'picture-books',
                'series-books',
                'young-adult',
                'young-adult-e-book',
                'young-adult-hardcover',
                'young-adult-paperback',
                'hardcover-graphic-books',
                'paperback-graphic-books',
                'manga',
                'combined-print-fiction',
                'combined-print-nonfiction',
                'animals',
                'business-books',
                'celebrities',
                'crime-and-punishment',
                'culture',
                'education',
                'espionage',
                'expeditions-disasters-and-adventures',
                'fashion-manners-and-customs',
                'food-and-fitness',
                'games-and-activities',
                'hardcover-business-books',
                'health',
                'humor',
                'indigenous-americans',
                'relationships',
                'paperback-business-books',
                'family',
                'hardcover-political-books',
                'race-and-civil-rights',
                'religion-spirituality-and-faith',
                'science',
                'sports',
                'travel',
                ]

current_time = datetime.now().strftime("%m%d%H%M%S")

# Overview of Best Sellers
query_str = base_url + overview + '.' + response_format + '?api-key=' +  api_key
```

```python
print(query_str)
request_results = requests.get(url = query_str)
data = request_results.json()

if 'message' in data.keys():

    print(data['message'])

else:

    for list in data['results']['lists']:

        list_name = list['list_name']
        display_name = list['display_name']
        list_books = list['books']
        header = [*list_books[0]]
        header.append('list_name')
        header.append('display_name')

        csv_dir = 'csv/' + current_time + '/best_sellers_overview/'

        if not os.path.exists(csv_dir):
            os.makedirs(csv_dir)

        csv_file = 'best_sellers_overview_' + list_name + '.csv'
        csv_output = csv_dir + csv_file
        with open(csv_output, 'w') as csvfile:

            # create handle to file
            writer = csv.writer(csvfile, delimiter=',')
            # write header
            writer.writerow(header)

            for book in list_books:

                # add list name to dictionary
                book['list_name'] = list_name
                book['display_name'] = display_name
                writer.writerow(book.values())

# Get Best Sellers in List
```

```python
for list_name in list_names:

    query_str = base_url + date + list_name + '.' + response_format + '?api-key=' +  api_key
    request_results = requests.get(url = query_str)
    data = request_results.json()

    print(data.keys())

    if 'message' in data.keys():
        print(data['message'])
        continue

    book_data = data['results']['books']
    header = [*book_data[0]]
    csv_dir = 'csv/' + current_time + '/best_seller_lists/'

    if not os.path.exists(csv_dir):
        os.makedirs(csv_dir)

    csv_file = 'best_sellers_' + list_name + '.csv'
    csv_output = csv_dir + csv_file

    with open(csv_output, 'w') as csvfile:

        # create handle to file
        writer = csv.writer(csvfile, delimiter=',')

        # write header
        writer.writerow(header)

        for book in book_data:

            # add list name to dictionary
            book['list_name'] = list_name
            writer.writerow(book.values())

    sleep(0.5)
```

8.2 R Code for Data Analytics

```
############################
# Import and Process Data
############################

library(corrplot)

data <-
read.csv(file='/home/matt/Documents/wpi/courses/s18/mis584/project/datasets/cafe/cafedata.csv'
, header=TRUE, sep=",")

data = data[5:22]
#########################################
# Scatter Plot and Correlation Matrix
#########################################

corrplot(cor(data), method = "circle")

########################
# Items Wasted Model
########################

lr_waste = lm(Total.Items.Wasted~. -Wraps.Waste -Total.Soda.and.Coffee, data)

summary(lr_waste)

########################
# Sales Model
########################

lr_sales = lm(Sales~.-Total.Soda.and.Coffee-Total.Items.Wasted,data)

summary(lr_sales)

########################
# Temperature Model
########################

lr_temp = lm(Max.Daily.Temperature..F.~.-Total.Soda.and.Coffee-Total.Items.Wasted,data)

summary(lr_temp)

lr_coffee = lm(Coffees~.-Total.Soda.and.Coffee -Total.Items.Wasted,data)

summary(lr_coffee)
```