DS 584 –Business Intelligence

Final Report on Real Foods (Hot Bar)

Submitted By:

Group 2

Siddhant Bhavsar, Jie Hu, Lyzanne Dsouza, Disha Kulkarni, Chevy Mac

Table of Contents

Table of Figures

1. Executive Summary

Our client company, 'Real Foods,' is a leading grocery store in the United States. It is an American multinational grocery chain known for its selection in organic food. It has more than 500 stores across the United States, with a workforce of around 91,000(Wikipedia, n.d.). The food business is one of the most profitable business sectors in the world. Real Foods recently had a merger with Amazon to become the biggest ever merger in the food sector. Considering this fact, one must not neglect the amount of food that is wasted. According to usda.gov, about 30-40% of the total food produced is lost, which is equivalent to 133 billion pounds of food every year. This contributes to $161 billion worth of food (Wikipedia, n.d.). One of the primary reasons responsible for wastage is the over-production of food.
Usually, cooked food has a short shelf life and should be consumed within a day or two. Real Foods is known for its cooked meals served at the 'Hot Bar' section of every store. 65% of sales at Real Foods generates from the 'Hot Bar' section. On average, Hot Bar contributes to 47% of the total profit generated at each store (Wikipedia, n.d.). We target on minimizing the food produced as well as increasing profit in this section by providing a BI dashboard for Real Foods store managers and executives.

1.1 Goals

We have data for 5 locations across three states in the Northeast United States. Locations are: Springfield, Worcester, Boston, Milford, and Providence
The goals for our project are:

1. Minimizing food wastage
2. Improve monetary benefits
3. Recommending food having low sales to wastage ratio

Our team started by summarizing the existing BI system of Real Foods. Our client is in contract with Amazon, which looks over the statistics of every store, but our client doesn't have an effective system to maintain records of items that have been produced (Wikipedia, nd). MS Excel is currently using for reporting and tracking. Amazon provides Real Foods AWS Cloud for better analytics. Our client uses the ERP system for groceries and supply chains. Moreover, they had a separate BI team as well, but that team didn't focus mainly on the Hot Bar section, they focused on products in the entire store. Our client doesn't use any BI tool or reporting system (WholeFoods, nd). Considering all this information, our client lies in the Teenager stage of the BI maturity model. Before starting to develop BI for Hot Bar section, we researched about competitors in a similar field. Our team found two major competitors: Leanpath and Phood Solutions. They both were building similar BI models, but it was complicated to use for employees working in the kitchen station of Real Foods. About 70% of the employee in the kitchen station of Real Foods speak Spanish and are not that fluent in English (Wikipedia, n.d.). Thus, our team planned to design a BI dashboard that would be easy to use for all.

Our team collected datasets from all these 5 locations. We had a dataset of around 65000 records for three months (October, November, December) and 22 columns. We pre-processed this data and cleaned it. Later, our team grouped the data based on dates to remove redundancy as there were multiple records of the same item name on the same day. Our team designed two dashboards.

1.2 Difficulties

While building these dashboards, we faced some difficulties:

1. Dataset had multiple records of the same item name on the same day but had different numeric values. Our team grouped those records to make it meaningful.
2. Negative values in the dataset: Weighing scale needs to recalibrate once the record has been generated. Failure in doing so would result in the generation of negative values in records. We removed such records from the dataset as they were giving inaccurate results.
3. Food Sector doesn't have many KPIs that we can explore unlike HR or Hospital sector, so our team had to come up with a new set of KPIs
4. We had a dataset of just 3 months, thus forecasting on such small DateTime range is difficult. Our team was not able to forecast for next month, but we built model to forecast sales and wastage for the next 2 weeks.

Dashboard1 provides information about the ongoing stats at Real Foods, such as the amount of food produced, sold, and wasted. Moreover, it also includes information about which is the top sold item and top wasted item. Dashboard being interactive, clicking on any location, or element would change other statistics and graphs to give more information. Our team also introduced the hierarchy feature in the dashboard for city and date attributes. One dedicated chart for daily sold and the wasted amount provided, which can be filtered as per manager's needs. All attributes in the dashboard were color-coded.

Dashboard2 had three major KPIs namely Over-production rate, % reduction in wastage and recovery rate (Donation)

Our team designed a butterfly chart for comparison of the top sold item to the top wasted item. Moreover, we created a line chart to determine the forecast of any selected item at any location. The item could be selected from the drop-down list—the sales vs. Wastage forecast model also created. The last and most crucial chart was the clustering chart, where we grouped top-performing items and worse performing items. Referring to this chart, managers can decide for which food item they should promote and which food item they should discard from their menu.

Moreover, if required for managers, our team built a model using SARIMA and Prophet modeling on Python to predict the wastage of food in the upcoming two weeks. Considering this statistic, managers can reduce the production of these items to prevent wastage and save money as well.

1.3 Project Output

Thus, benefits for Real Foods by using our designed dashboards are:

1. Our dashboards will help Real Foods in reducing wastage by ~30 % by providing significant insights about our items.
2. It can quickly gather visual data based on the manager's needs.
3. Adding predictive analytics, as shown in dashboards, can help Real foods predicting the profitability of items in their menu.
4. Overall using the dashboards will optimize the tiring process of selecting which products to keep or discard in the next month from their list.

2. Introduction

The client company for our project is called "Real Foods Market" Inc. Real foods Inc. is a multinational grocery chain with its headquarters in Austin, Texas, and its parent company being Amazon. The company was founded 39 years ago on September 30th by John Mackey, Renee Hardy- Lawson, Mark Skiles, and Craig Weller (Wikipedia, nd). It primarily offers organic and free from hydrogenated fats, preservatives foods (Wikipedia, nd). Along with organic grocery, the company is a seller of houseware, wine, and other items as well. In the grocery stores, the company offers a wide range of categories of items such as Dairy and eggs, Bakery, Seafood, etc. and foods by dietary preference such as Vegan, Gluten-free, Keto-friendly, Sugar Conscious, low sodium, and many more. Hence, the company falls into the grocery store and the health food industry (Wikipedia, n.d.).

The company has more than 500 stores in the United States. As of today, the total strength of the company is 91,000 employees. In terms of management, "Real Foods" operates in twelve geographic regions, and each of these regions has its own President, regional administrative team, store-level leadership, and store-level team members. Also, the company functions on a four-tier hierarchy of employment, such as Store Employment, Facilities Employment, Regional Offices, and Global Headquarters (Wikipedia, n.d.).

For our project, we have chosen to focus on the "Prepared Foods" severed at the Hot Bar section of "Real Foods". Prepared foods are any item that is cooked fresh daily and served daily to the customers. Some of the items prepared and served in the hot bar section are Mac and cheese, Chicken tenders, steamed broccoli, Mashed Potatoes, and many more. 65 % of the sales come from prepared foods at the hot bar. Since the food is cooked daily and served daily, more than 30000 lbs. is wasted per year per location. This creates a sense of urgency for the company. For our project, we have considered five Real Foods location, that is Springfield, Worcester, Boston in Massachusetts, Providence in Rhode Island, and Milford in Connecticut. Our focus for this project has been these five locations because we had the data for these five locations. Noticing the enormous amounts of data that is wasted daily at the Hot bar section, our team decided to work on providing solutions for using the Business Intelligence tool, Tableau. The goal of our project is to analyze food wastage at the Hot bar section and provide them with our recommendations to curb the same.

Currently, the team at the hot bar section do have any Business Intelligence tool for reporting and analysis purpose. They are using Microsoft Excel to perform their analysis and reporting work. The users of MS excel are the managers at Real Foods hot bar department. The department has a separate BI team, but it does not use the BI reporting tool to be performing its analysis and reporting, and they do not focus on hot bar wastage. Overall, the company has data warehouses and uses a "single-view" ERP system for managing its data. This happened in 2013 when "Real foods" dug into its in-house technology to determine whether it could support future challenges. The company realized that approximately 90% of its IT infrastructure was filled with obsolete legacy systems, and it needed to be replaced with the current technology. Hence, to get rid of its outdated legacy systems, "Real Foods" had a merger with a company named "Infor" which is a leader in ERP vendors amongst other competitors such as Oracle, Microsoft, SAP. Infor helped the company by providing "the next generation ERP system for retail." Real foods have their reasons to transition from old-fashioned legacy systems to ERP, one of them being if one change had to be performed at one instance per location, it had to be done at all other locations for those number of times. Thus, it was tedious. With the new single view ERP system for retail, the

company can make the change simultaneously on all the locations at once replacing the old approach of doing it "n" number of times for "n" number of locations. Also, the legacy system cannot accept change in the way business functions. Thus "Real foods" needed to replace its legacy system. This change in technology was estimated to make the US $300 million savings due to the efficiency of the ERP. Thus, "Real foods" expect the ERP to make huge savings, modify the way the company does business as cited by the CIO in 2016, be more efficient, and abolish the deficiencies of the existing antique legacy systems. Later, in the year 2017 Amazon had a 13.7 billion acquisition of "Real Foods" grocery chain, making it the current parent company of the grocery store. ERP Vendor, "Infor" uses AWS as a third-party platform for providing cloud solutions to its clients instead of developing its infrastructure, thus making "Real Foods" a user of AWS (Amazon Web Services). Thus, all the issues related to "Real foods" supply chain and retail are solved using Cloud and data analytical solutions from AWS. However, this is not the case with the hot bar section at "Real Foods". They have a BI team, but it does not concentrate on food wastage, it focusses on the overall Real Foods grocery store. This is the reason; our team feels that the hot bar section is on the "Teenager" Stage and by using our proposed BI solutions, we hope it would transition into the adult stage as shown in fig 1.

BI Maturity Model

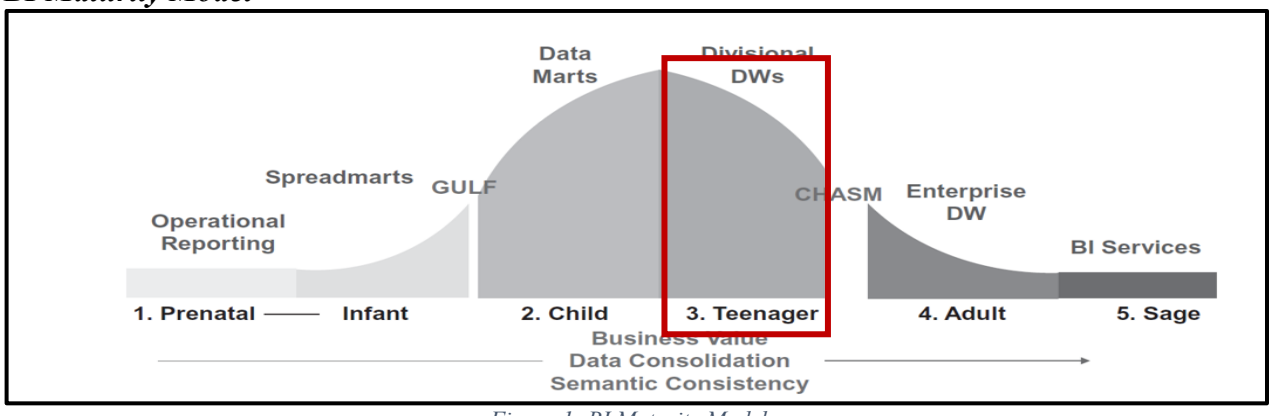


Figure 1: BI Maturity Model

Following are the reasons why the company should adopt BI solutions and ways in which BI can help them -

- Using predictive analysis and forecasting features available in our dashboard, the managers can learn about the future and set their targets accordingly. Also, to do this, one must not be a high-end coder as this can be done through tools like BI Tableau.
- It will allow the managers of the hot bar section to understand it's "top sold" and "top wasted" items, knowing this they can focus on these top products and understand their needs. The company can leverage this to understand customer choices, produce accordingly, and avoid wastage. Also, by knowing customer choices, the company can advertise certain products and make targeted offers.
- Lastly, BI will also help them understand which item is most profitable and which is not, this will make them focus on the most profitable ones, understand why some products are not profitable and either improve those item's selling strategy or discard them completely as keeping them is causing more costs than profits.

2.1 Our Users

The kitchen staff who cook the food and the managers who make the decisions regarding the food production would be the primary users of our recommended dashboards as they are the ones who can make critical decisions about the food wastage.

2.2 Case Studies

For two case studies, which were Lean path and Phood solutions, both utilize to prevent food wastage. Our team researched their approach to treat food wastage.

2.2.1 Leanpath

We studied and analyzed two case studies to better design our dashboards. The following paragraphs will elaborate on each of the case studies.

The lean path creates food waste prevention IT solutions. These solutions educate staff about the environmental and financial impact of the food wastage. The data related to the food item is automatically uploaded on the lean path online website, which is a cloud computing website. Lean path helps in weighing the food and tracking its wastage, giving the complete user visibility on how much wastage is being generated (Leanpath Food Waste Prevention Solution, n.d.). The lean path is a strong believer of the fact that staff working in the kitchen are the reformers of food wastage and can influence others with similar wastage problems.

Lean path is a strong believer of the fact that staff working in the kitchen are the reformers of food wastage and can influence others with similar wastage problems (Leanpath Food Waste Prevention Solution, n.d.). Leanpath setup looks like in fig 2.

Figure 2:Leanpath Setup

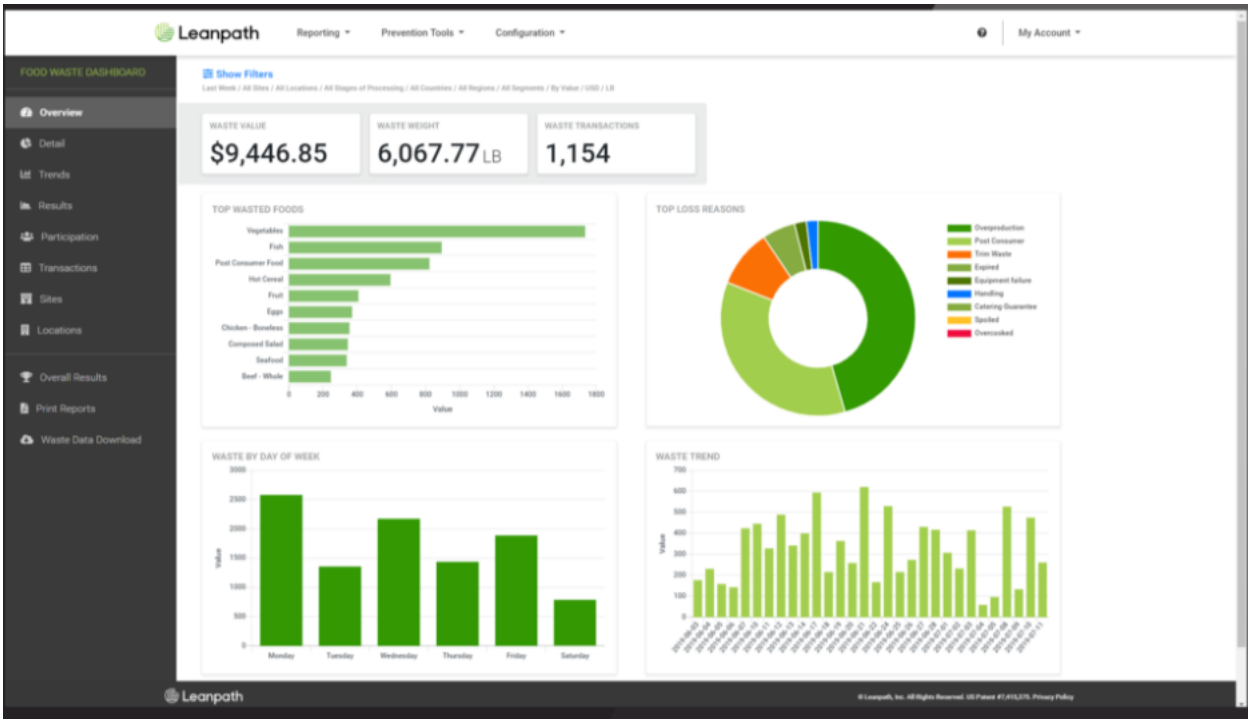


Figure 3: Leanpath Dashboard

Leanpath dashboard comprises of donut chart, and couple of bar charts. Leanpath doesn't provide any analytical tool for their dashboard. Moreover, they only provide operational dashboard.

2.2.2 Phood Solutions

The Phood Solutions targets on the reduce food wastage market. Their clients include multiple educational institutes such as Tu Delft University, University of Minnesota, University of Cincinnati, etc. They recently collaborated with grocery chain giant Whole Foods, which is currently their primary client (Food Waste Ends Here, n.d.). The Phood Solution provides simple, easy-to-use software and hardware, which helps businesses save over $100,000 per year by reducing 50% of their food waste, as well as increase excess food donations by 35% (Food Waste Ends Here, n.d.). The Phood solutions have a straightforward, inexpensive waste calculation systems which measure the exact food items wasted on a day-to-day basis, providing valuable insights about loss reasons and staff behavior. These metrics give their client the power to alter wasteful food purchasing, ineffective food preparation, and storage, and create a kitchen operating at max efficiency. They provide four different types of hardware setup for recording food wastage (Food Waste Ends Here, n.d.). The client can choose one of these based on their needs. Once the hardware has set up, the employee has to weigh their product in a tray and then place it on the hot bar. Every time the food item goes out from the kitchen, the record has generated.

Similarly, if the item is discarded or donated in the end, still the record is generated. Due to this, the same food item might have multiple records on the same day. Based on these records, the Phood solution provides a dashboard to their manager to keep track of their daily food production and wastage (Food Waste Ends Here, n.d.).

Figure 4: Phood Solutions Dashboard

Their dashboard hardly provides any significant information about the food that is being most wasted, most sold, etc. Moreover, the charts are not uniformly color-coded. The Phood solutions use D3 JavaScript to design their dashboard. It has compatibility issues with lots of models. We plan on using the Phood solutions approach in collecting data and building a new dashboard that will provide insightful information to the store managers. Moreover, our team will also develop an analytical dashboard to recommend store managers about their top-performing food items and worse performing food items.

3. The Proposed Business Intelligence Solution

Business intelligence (BI) leverages software and services to transform data into actionable insights that inform an organization's strategic and tactical business decisions. BI tools access and analyze data sets and present analytical findings in reports, summaries, dashboards, graphs, charts, and maps to provide users with detailed intelligence about the state of the business (Pratt, M., & Fruhlinger, J. (2019, October 16)).

3.1 BI Components

The major components of Business Intelligence Solutions are as below:

- **Online Analytical Processing (OLAP):** OLAP analyses business information in a multi-dimensional manner to assist with complex calculations, trend analysis, and data modeling. With OLAP, the end-user gets an opportunity to analyze specific data in multiple dimensions to obtain the necessary insight for making a decision.

- **Corporate Performance Management:** Corporate Performance Management (CPM) encompasses methods, metrics, activities, and systems, which have used to monitor and manage the business performance of a company. CPM software processes the focused information to turn it into operational plans.

This process and methodology offer business owners an integrated approach to planning, forecasting for finance, sales, marketing, HR, and operations. When this methodology has implemented, it joins company strategies with plans and executions, thus helping a business succeed and improve.

CPM is an essential component of business intelligence for companies that are looking for such changes as budget remodeling, cost-cutting, upgrading organization strategy, better KPIs alignment, and improving the process of financial planning.

- **Real-time Business Intelligence**: Real-time business intelligence (RTBI) has used when sorting and analyzing business data and operations have to be done at the collection stage. Real-time BI allows the company to get insights into the business process as quickly as possible to take strategic action.

- **Data Warehousing**: Data warehousing allows the business owner to go through different data subsets and examine components that could help make the right business decisions. It helps create important statistics about the business and the industry. Warehousing implies storing formidable amounts of data in numerous unique ways, which could be useful for analysis.

- **Data Sources:** For the business intelligence process and methodology to be integrated, it is crucial to have the right understanding of the data sources. Pulling raw data from different sources, internal and external, is vital to the diverse analysis options.

  Companies tend to store vast amounts of operational data. BI needs to navigate between the data sources. In most companies, mainframe legacy systems create a foundation for the data centers because they can deal with large volumes of data. However, such data is usually challenging to procure since many legacy apps are often obsolete or proprietary. Operational data sources to take advantage of our Enterprise Resource Planning, CRM (Customer Relationship Management), E-Commerce Apps (Understanding the Major Components of Business Intelligence. (2019, August 30)).

3.2 Our proposed Business Intelligence Solutions

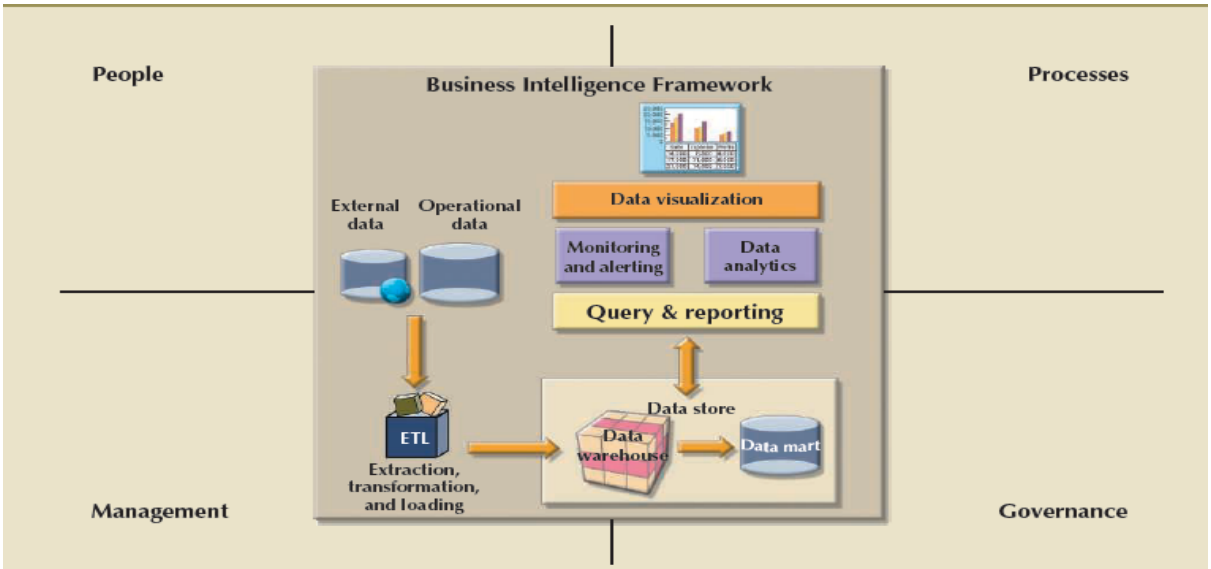


Figure 5: Business Intelligence Framework

- **External and Operational Data:** Considering only the Hot Bar section at Real Foods, the data sources could include daily operational data related to sales, wastes, inventory, i.e. production collected from every outlet/location individually. This data is manually fed into the data systems by the daily operational workers at the Hot Bar kitchens at Real Foods.

- **Extraction, Transformation, and Loading:** The data from the operational databases at each location/outlet can then be loaded into a data warehouse at a daily refresh rate. It can be done using BI Tools like Tableau that allow users to set the load and refresh rates as per their choice. It will also help Real Foods maintain consistency in their data at all stores and gain better value from the data.

- **Data Warehouse and Data Marts:** The below star schema is our proposed data warehouse model recommended for Real Foods (Hot Bar Section):

3.2.1 Star Schema for Hot Bar – Real Foods

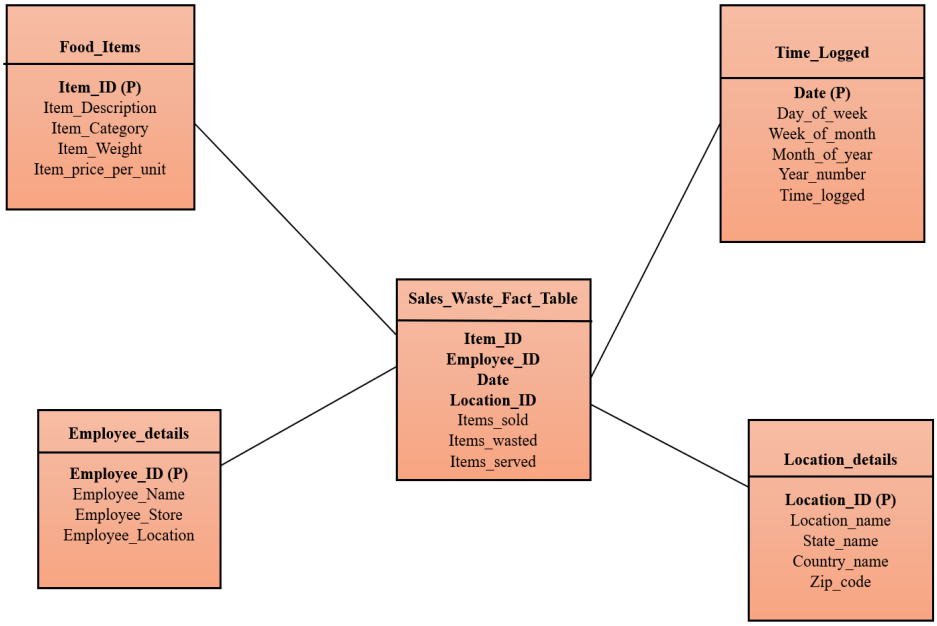


Figure 6: Star Schema

The above star schema includes 4 Primary tables and 1 Fact table. The Primary tables include Food Item details, Employee Details, Location Details, and Time Logged Details. The Fact table is based on the sales and wastes details co-dependent on other primary tables.

1. **Food Item Details:** This table is used to store data related to the food items being served at the Hot Bar Section. It includes the item description, weight, category, and price.

2. **Employee Details:** This table is used to store data of which employee has logged the data w.r.t the food served, sold, and wasted at different outlets.

3. **Logged Time Details:** This table is used to store the time and date details for the food items recorded by the employees at various Real Food outlets.

4. **Location Details:** This table is used to store the data related to the location of Real Foods stores where the food items are either sold or wasted.

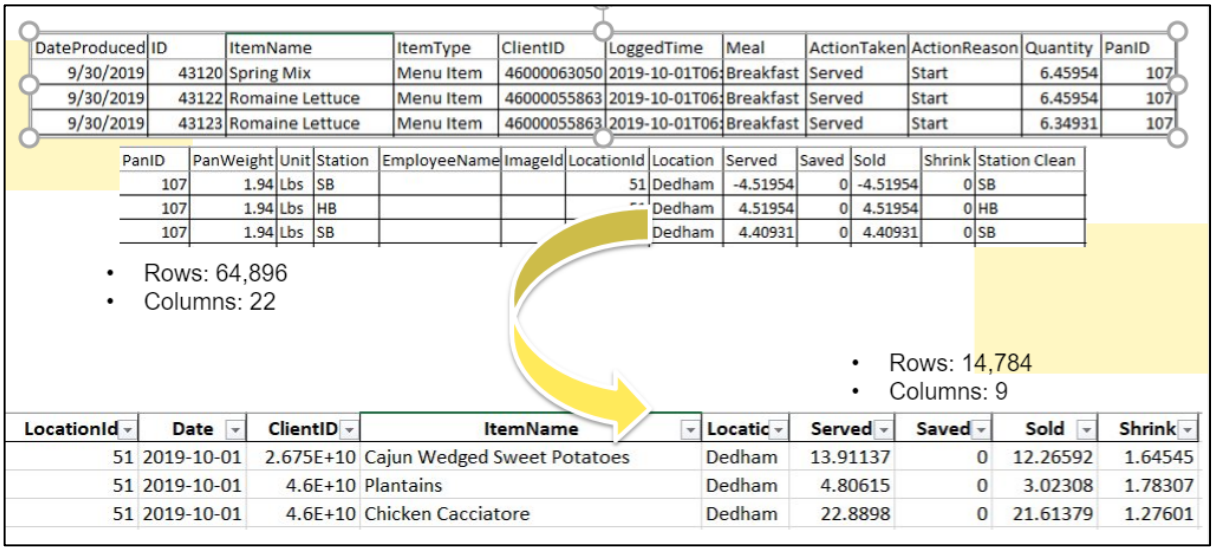
- **Monitoring and alerting:** In our proposed BI solution for the Hot Bar section, we recommend certain charts using BI tools like Tableau. It would enable the users to set a Waste Limit Cut-off that would help the store and regional managers monitor which store outlets exceed the waste limit set. Also, some charts would help the managers monitor the weekly and daily trends for sales and wastage of food items.

- **Data Analytics:** For Data Analytics, we recommend using BI tools like Tableau in which we have created forecast time series models that help predict the future trends for the wastage and sales. It will help the managers to foresee the overproduction of certain food items and could probably adjust their production rates accordingly. We have also used Python to build predictive forecast models like SARIMA and Prophet to get the forecast of the next 15 days for the top 10 wasted products. Using these models, the end users can take calculative decisions and prevent wastage and loss of food items accordingly.

- **Data Visualization:** For Data Visualization, we recommend using the two Dashboards that we have developed using Tableau. These dashboards will help the store and regional managers to get insights on how much sales and wastage each outlet is contributing to, as well as which items are performing well or not. The dashboards will also give the managers an analytical view of forecasts and trends for various food items. Many managerial questions can be answered using these dashboards, which will indeed help them make better decisions.

4. Three Use Cases/Prototypes

The **data set** we used comprises operational data for the period of three months (October 2019 – December 2019) for five outlets. The five outlets include Boston, Springfield, Worcester, Milford, and Providence spread across three states. We specifically focus on providing BI solutions for the Hot bar section, and hence we cleaned our data and pre-processed it to encompass data only related to this section. For data cleaning and pre-processing, we used tools like Python and Excel. The initial dataset had 64,896 rows and 22 columns. However, to only be specific to the Hot Bar section and on performing data cleaning, we have 14,784 rows and nine columns.



DateProduced	ID	ItemName	ItemType	ClientID	LoggedTime	Meal	ActionTaken	ActionReason	Quantity	PanID
9/30/2019	43120	Spring Mix	Menu Item	46000063050	2019-10-01T06:	Breakfast	Served	Start	6.45954	107
9/30/2019	43122	Romaine Lettuce	Menu Item	46000055863	2019-10-01T06:	Breakfast	Served	Start	6.45954	107
9/30/2019	43123	Romaine Lettuce	Menu Item	46000055863	2019-10-01T06:	Breakfast	Served	Start	6.34931	107

PanID	PanWeight	Unit	Station	EmployeeName	ImageId	LocationId	Location	Served	Saved	Sold	Shrink	Station Clean
107	1.94	Lbs	SB			51	Dedham	-4.51954	0	-4.51954	0	SB
107	1.94	Lbs	HB				Dedham	4.51954	0	4.51954	0	HB
107	1.94	Lbs	SB				Dedham	4.40931	0	4.40931	0	SB

LocationId	Date	ClientID	ItemName	Locatic	Served	Saved	Sold	Shrink
51	2019-10-01	2.675E+10	Cajun Wedged Sweet Potatoes	Dedham	13.91137	0	12.26592	1.64545
51	2019-10-01	4.6E+10	Plantains	Dedham	4.80615	0	3.02308	1.78307
51	2019-10-01	4.6E+10	Chicken Cacciatore	Dedham	22.8898	0	21.61379	1.27601

Figure 7: Dataset wrangling

Below are our proposed Business Intelligence solutions in detail.

4.1 Real Foods Dashboard1

Our **first proposed prototype**/dashboard is as below. This dashboard will help store managers and regional managers to view metrics as Units Sold, Units Wasted, and Units Served at different locations and outlets. This dashboard will also help them to get insights on the top-selling food items and the top wasted food items along with their weekly trends.

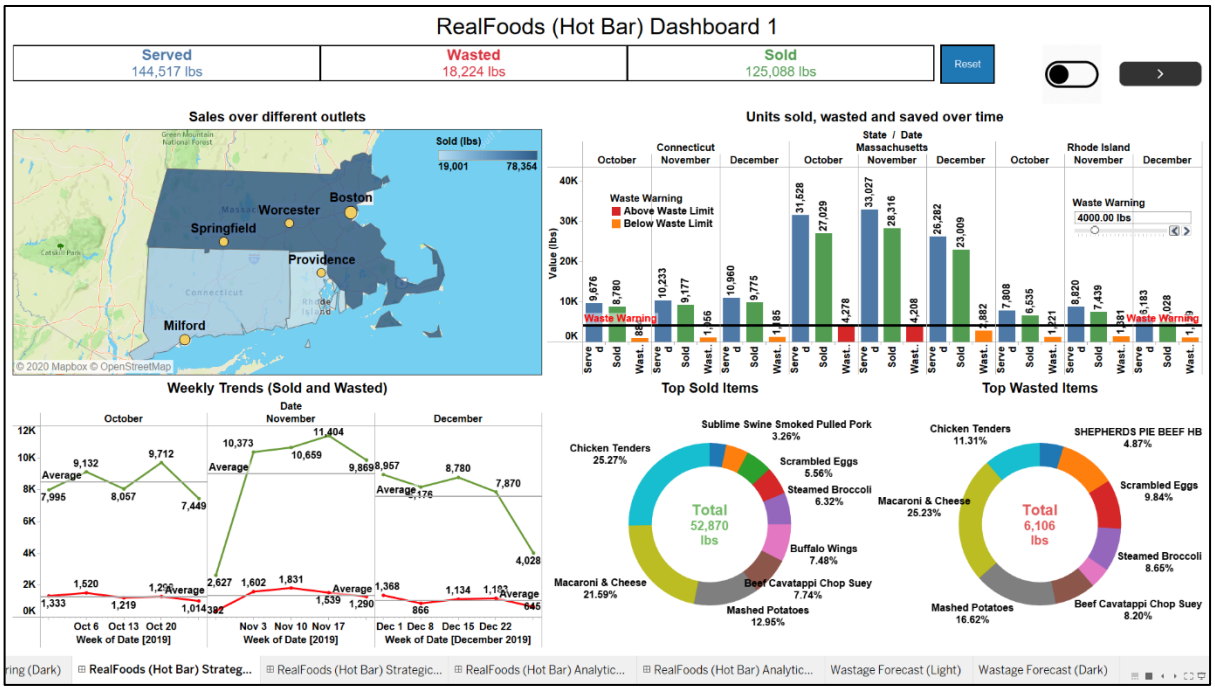


Figure 8: Real Foods Dashboard1

We have used the below graphs and features for Dashboard1:

1. The first section of our Dashboard1 comprises of the metrics (Served, Wasted, and Sold) in-lbs, which is the sum of the metrics at all five locations.

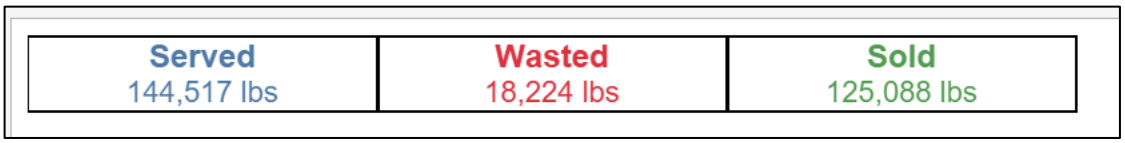


Figure 9: Real Foods Dashboard1 KPIs

2. The first graph is a map used to show the sum of sales at different outlets. We have used a color filter to enhance the visualization and portray the number of sales at both state and city level.

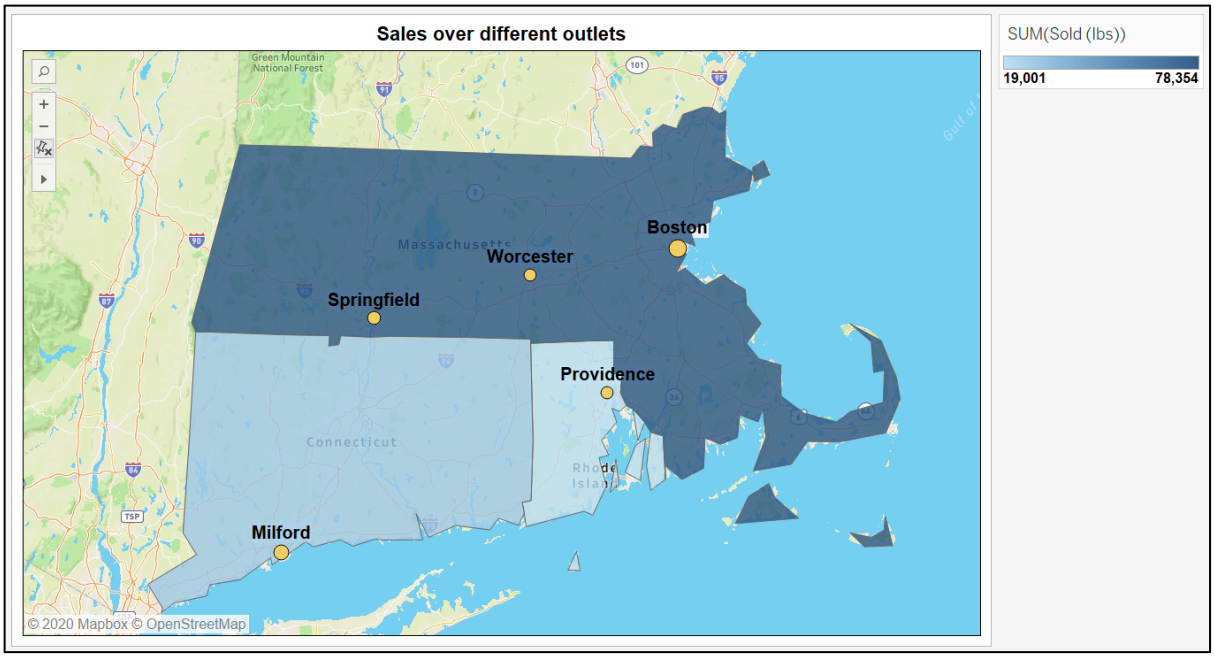


Figure 10: Real Foods Dashboard1- Sales over Outlet

3. The third graph is a bar graph that displays the units sold, wasted, and saved over a period of three months. In order to get better insights into the data, we have also added a hierarchy to the location and date dimensions. Hence, we can drill down the location as State and City and Date as Month and Day.

   The graph also consists of a parameter named Waste Warning Parameter. This parameter enables the end-users to set a waste limit of one's choice and monitor if a particular location has exceeded the waste limit set for a particular period of time.

   For example, in the below screen, we have set the Waste Warning Parameter as 4000 lbs, and it can be seen that for October and December in Massachusetts, the limit was exceeded, and hence it was marked red to indicate an alert.

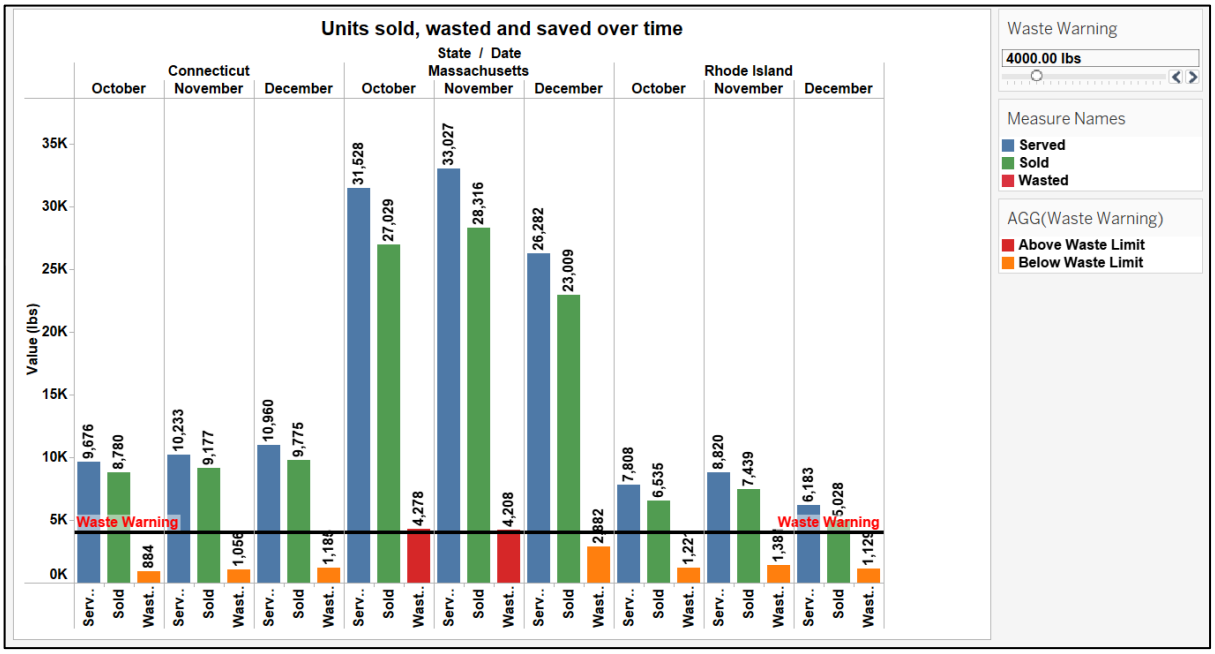


Figure 11: Units sold, wasted and saved over time for Dashboard1 with hierarchy-1

The below screen displays the hierarchy, State and City for the units sold, wasted, and saved in-lbs along with the Waste Warning Parameter set at 2000 lbs.

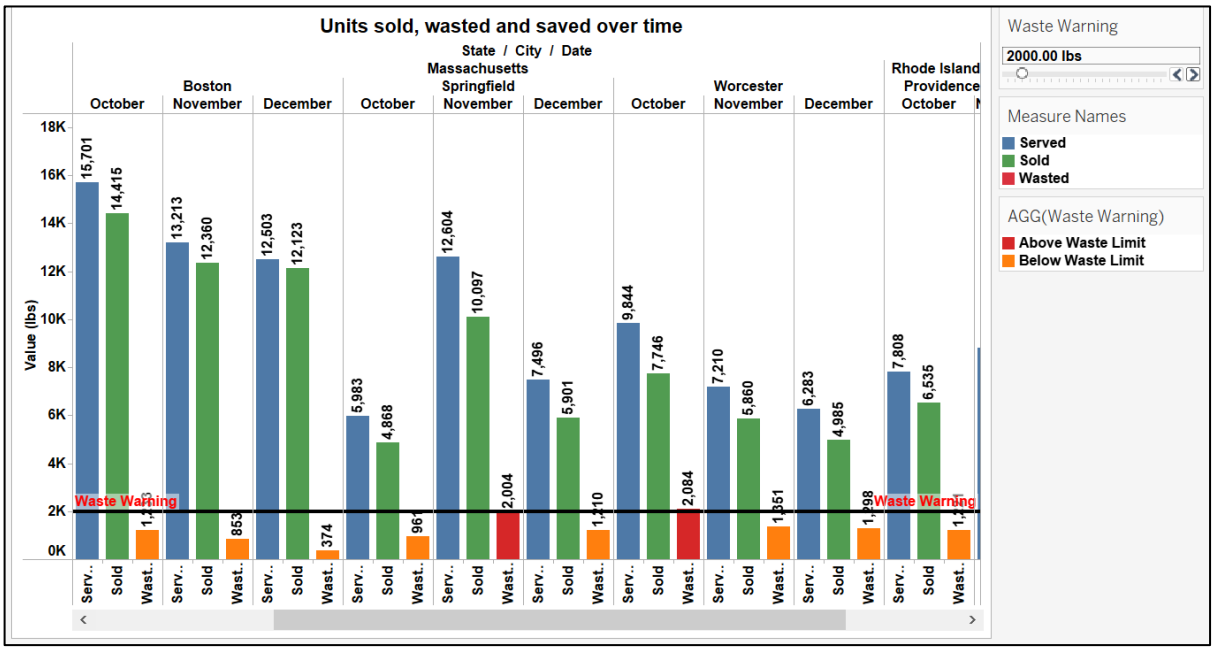


Figure 12: Units sold, wasted and saved over time for Dashboard1 with hierarchy-2

4. The fourth graph is a weekly trend graph that displays the weekly sales and waste trends over a period of time. We have also added a hierarchy for Date, which enables the end-users to see the trends on a weekly as well as a daily basis for both sales and wastage.

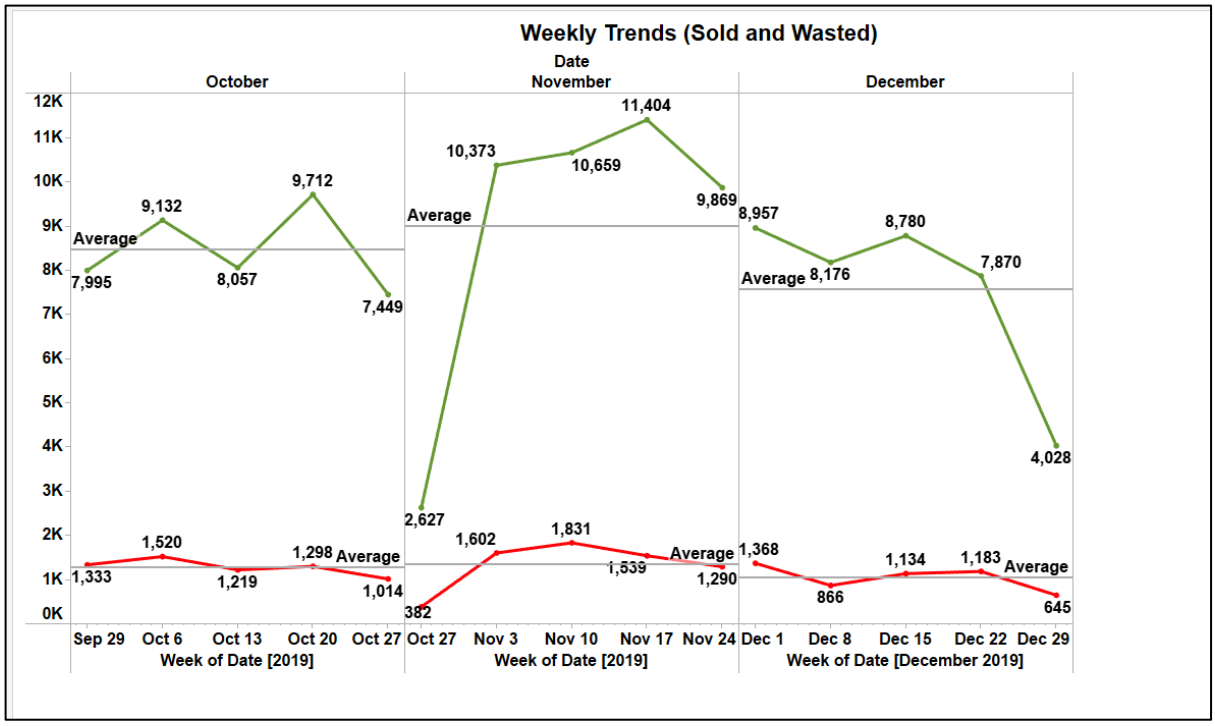


Figure 13: Weekly Trends (Sold and Wasted) time Hierarchy-1

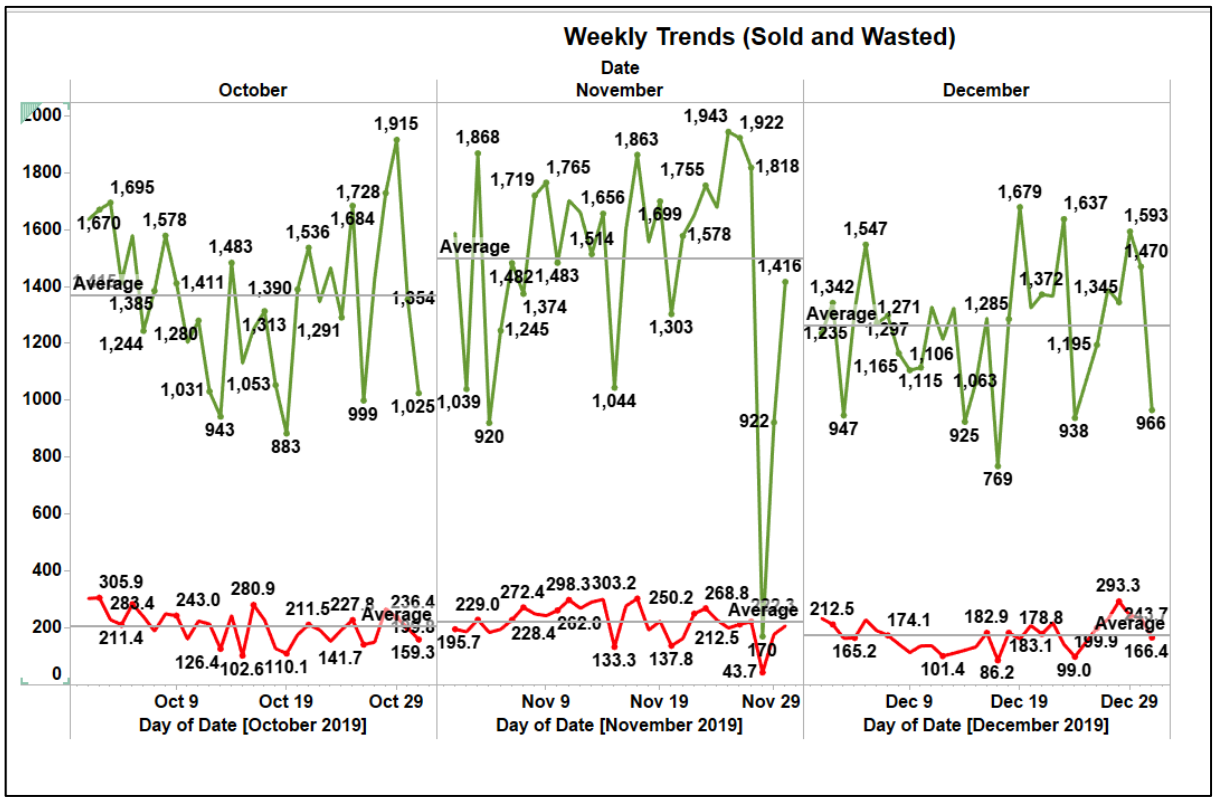


Figure 14: Weekly Trends (Sold and Wasted) time Hierarchy-2

5. The fifth graph is a donut chart that is used to display top sold items. Here, we have considered the sum of sales for selecting the top 10 items. This graph enables the end-users to get a better understanding of which food items are performing well.

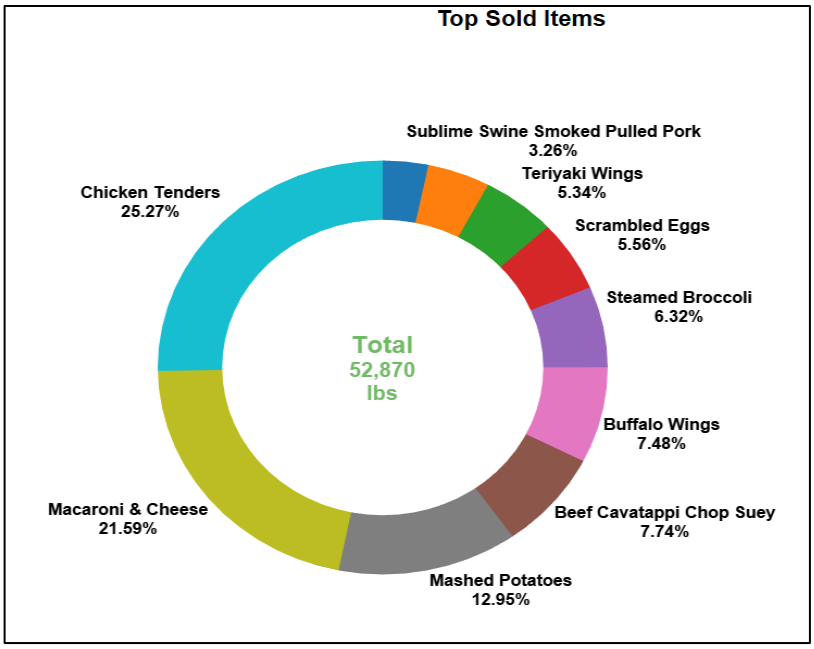


Figure 15: Top 10 Sold Items Dashboard1

6. The sixth graph is a donut chart that is used to display top wasted items. Here, we have considered the sum of wastes for selecting the top 10 items. This graph enables the end-users to get a better understanding of which food items are not performing well.

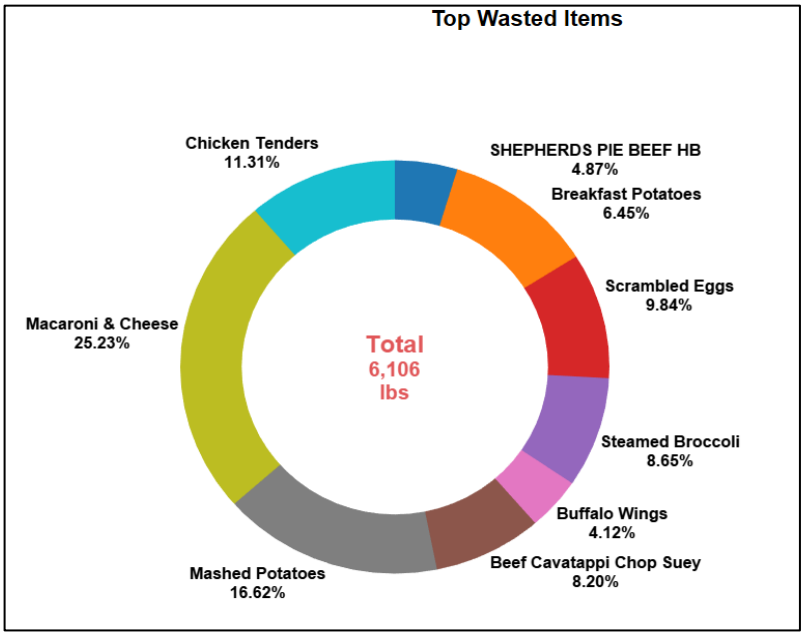


Figure 16: Top 10 wasted Item Dashboard2

7. For making the Dashboard1 more interactive and user-friendly, we have added several interactive and visualization features and effects.

    - We have added an action filter on the different sheets of the dashboard, which helps the end-users to see the correlation and visualize data more intensively.

    - We have also added a Reset, Navigate, and Theme Button. The reset button helps refresh the dashboard to normal if any filter was applied to it earlier by the user. The Navigate button helps the users to navigate to the second dashboard. The Theme button has been developed as an optional feature to give users the benefit of viewing the dashboard in dark mode.

    - Below are the screens that display the features and effects added to Dasboard1:

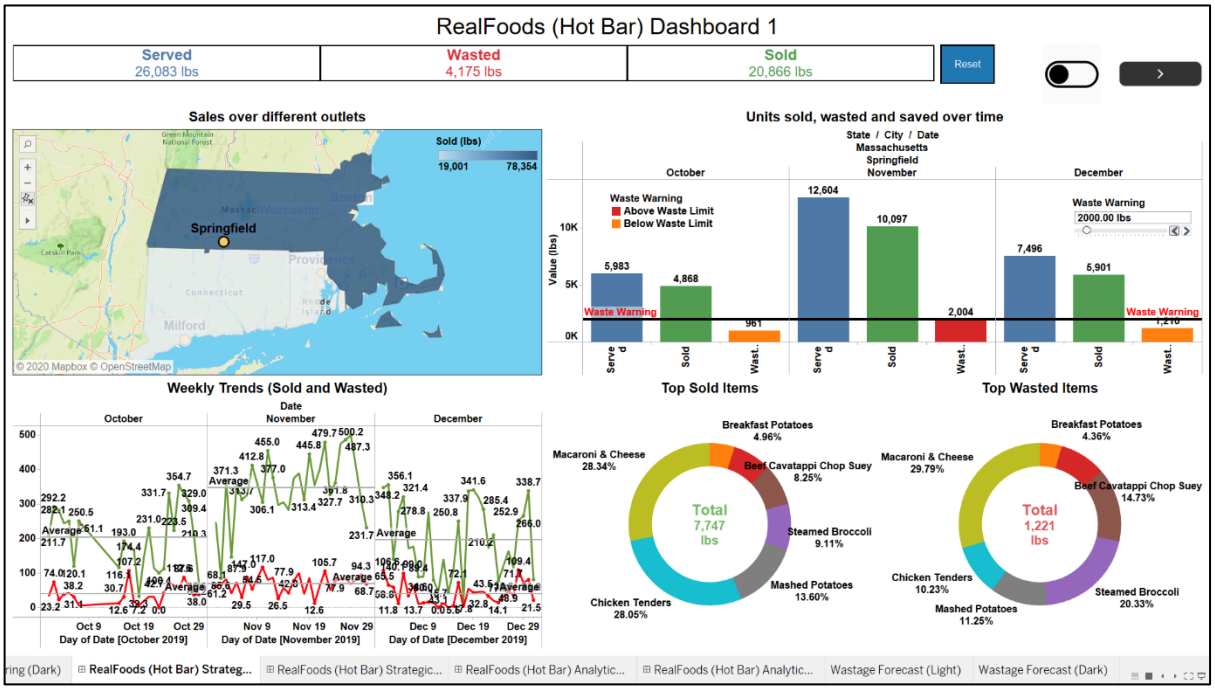


Figure 17: Real Foods Dashboard1

Here, in the example above, we have selected Springfield, MA, in the graph and can see that the action filter enables the user to see data only related to Springfield in other graphs as well.

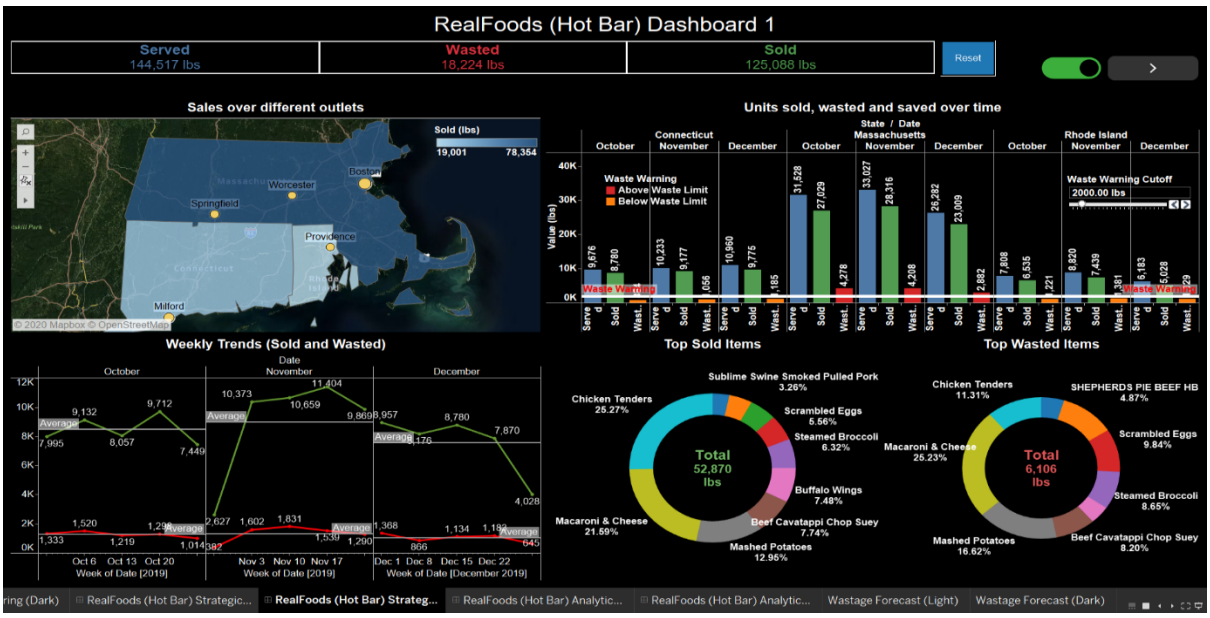


Figure 18: Real Foods Dashboard 1 Black Theme

Here, in the graph above, we have selected the Dark Mode of the Dashboard1, which is an optional feature for the end-users.

Dashboard1 will help store and regional managers answer the below managerial questions:

- Which outlet/location has the most sales?
- What are the top-selling food items?
- What are the top wasted food items?
- What is the average weekly and daily forecast of sales and waste?
- Which locations exceed the waste limit for a particular time period?

4.2 Real Foods Dashboard2

Our **second proposed prototype**/dashboard is as below. This dashboard will help regional and store managers to get better insights on the future forecasts and trends of the food items, sold and wasted at different locations using the different charts on the dashboard.

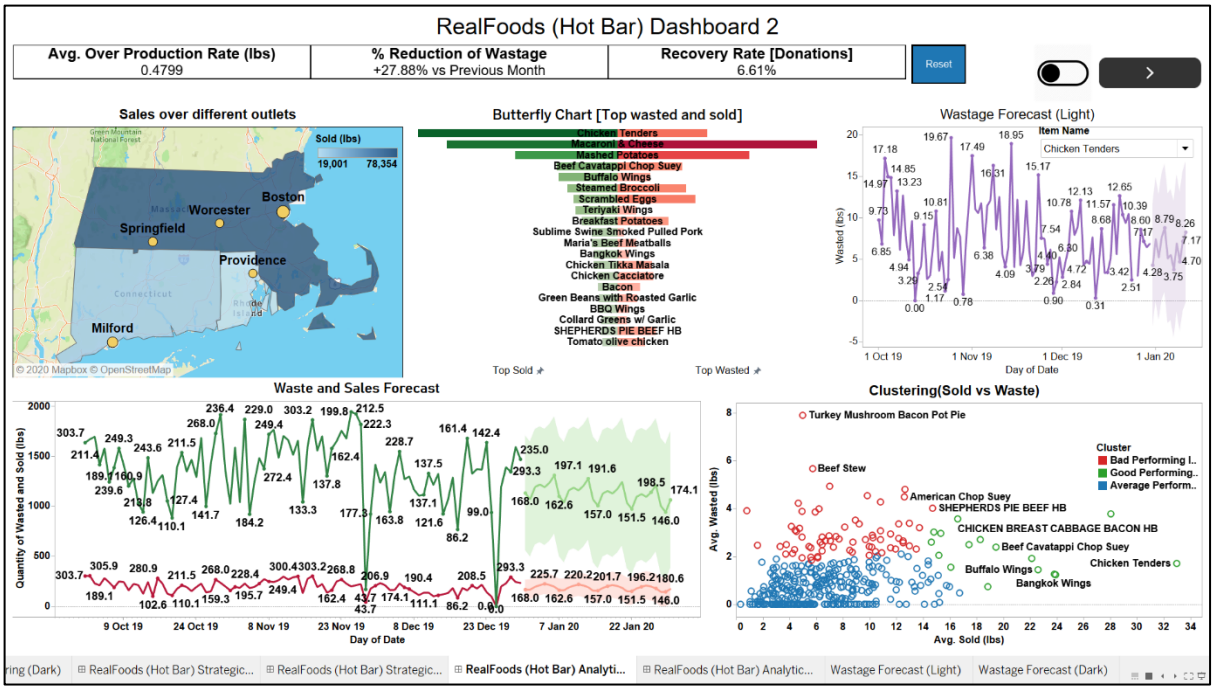


Figure 19: Real Foods Dashboard 2

We have used the below graphs and features for Dashboard2:

1. The first section on our Dashboard2 is the Key Performance Indicators. We have considered three main KPIs, namely, Average Over Production Rate, Percent Reduction of Wastage, and Recovery Rate.

    - **Average Over Production Rate:** The average over production rate is calculated by dividing the sum of wasted units by the sum of sold units. It helps to estimate the number of units that are overproduced over a period of time.

    - **Percent Reduction of Wastage:** The percent reduction of wastage is calculated by subtracting the sum of wasted units of the current month from the sum of wasted units of the previous month and further dividing it by the sum of wasted units of the current month. (SUM ([Wasted PM])-SUM ([Wasted CM]))/SUM ([Wasted CM])

      The goal here is to monitor the percent reduction wastage every month and strive to reduce it in order to avoid wastage in the future.

    - **Recovery Rate:** The recovery rate is a KPI that can help the end-users if they are interested in donating the wasted or remaining food to particular food banks. It is calculated by dividing the sum of donated units by the sum of wasted units. The higher the recovery rate, the better the overall donations made by the company.

Avg. Over Production Rate (lbs)	% Reduction of Wastage	Recovery Rate [Donations]
0.4799	+27.88% vs Previous Month	6.61%

2. The first graph is a map that is used to show the sum of sales at different outlets. We have used a color filter to enhance the visualization and portray the sum of sales at both state and city level.

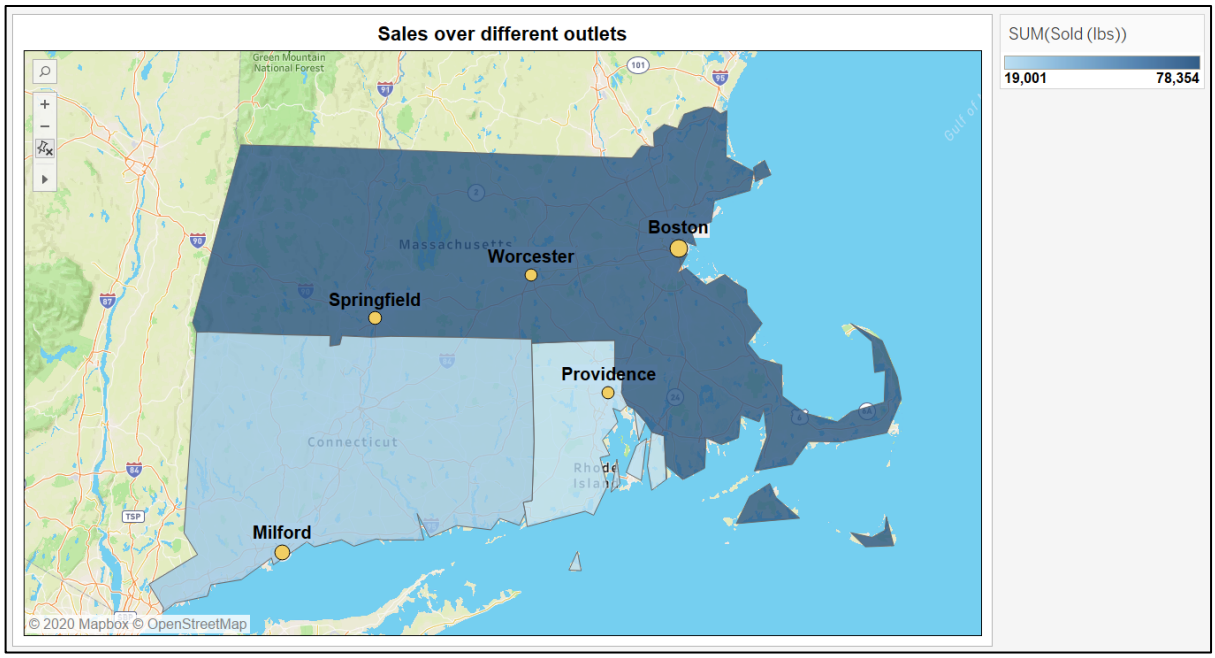


Figure 20: Sales over Different Outlet Dashboard2

3. The second graph is a butterfly graph which displays the top sold and wasted items. It helps the end-users get an understanding of the items performing well and not performing well in contrast.

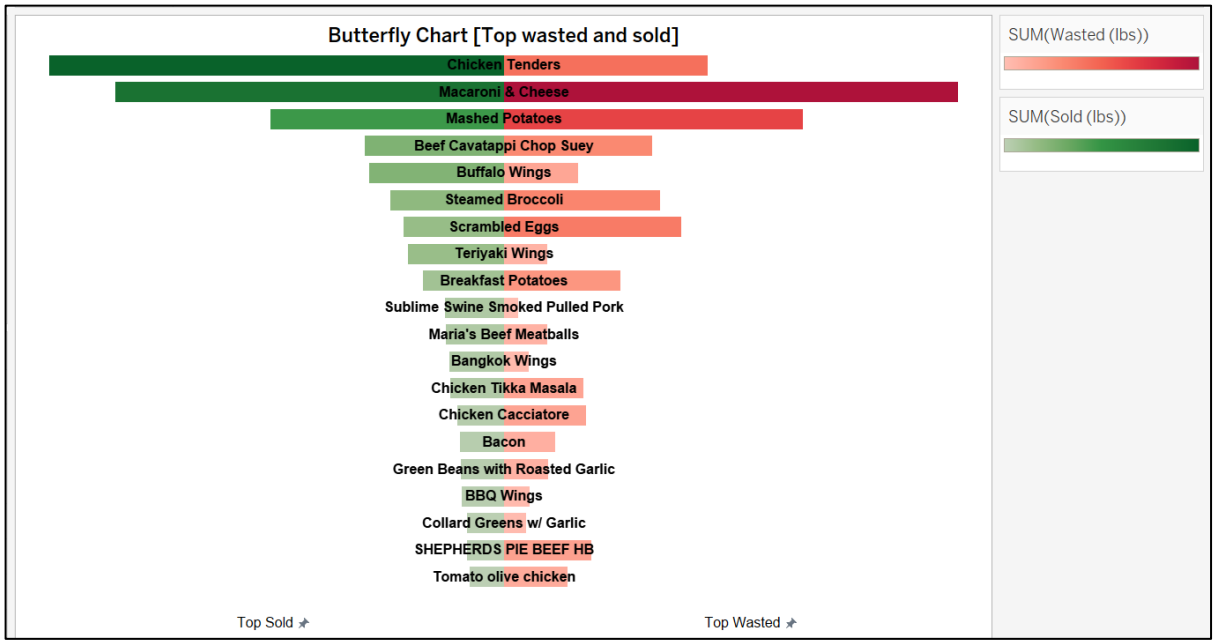


Figure 21: Dashboard2 butterfly chart

4. The third graph is a wastage forecast graph for food items and predicts the waste for the next month. We have added a single drop-down list that helps the end-users to select an item of their choice that they want to see the forecast for. The below example shows us the screen where we have set the item as Chicken tenders and displays the forecast for the same. We have used a custom forecast model with additive trends and additive seasoning per day.

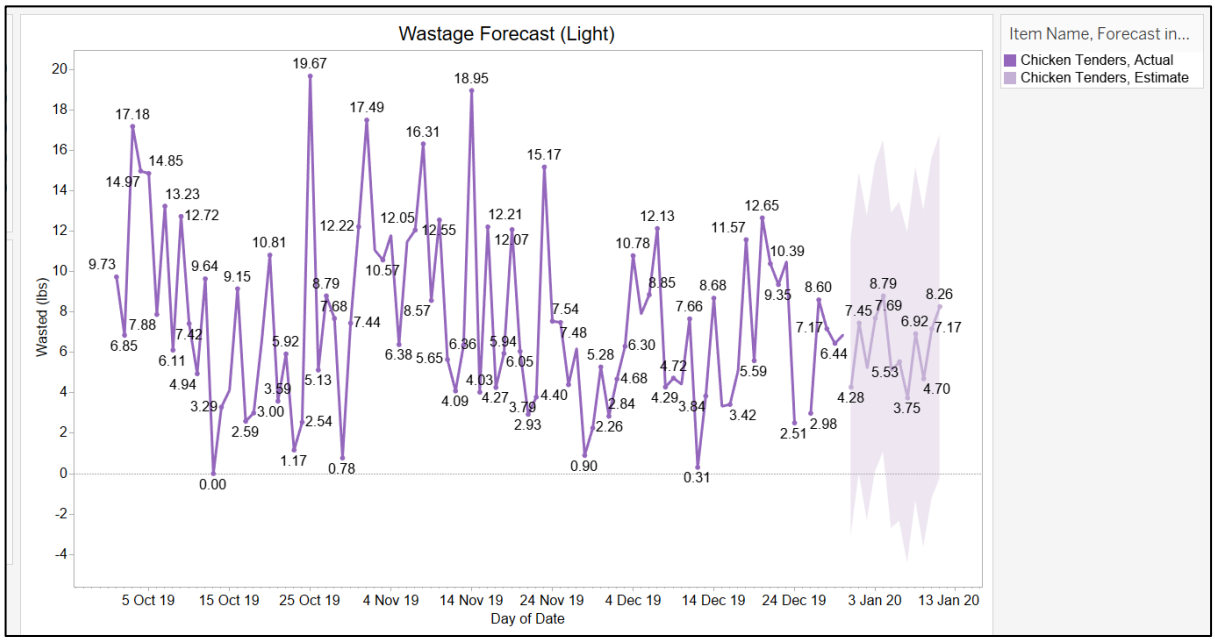


Figure 22: Dashboard2 Sales Forecast

5. The fourth graph is the sales and wastage forecast graph. For this graph, we have added a custom forecast model that has additive trends and additive seasoning per day. This forecast helps the users analyze and get a better forecast of how their sales or wasted units stand for the next month. They can accordingly modify or make changes in their products to help decrease waste and improve sales accordingly.

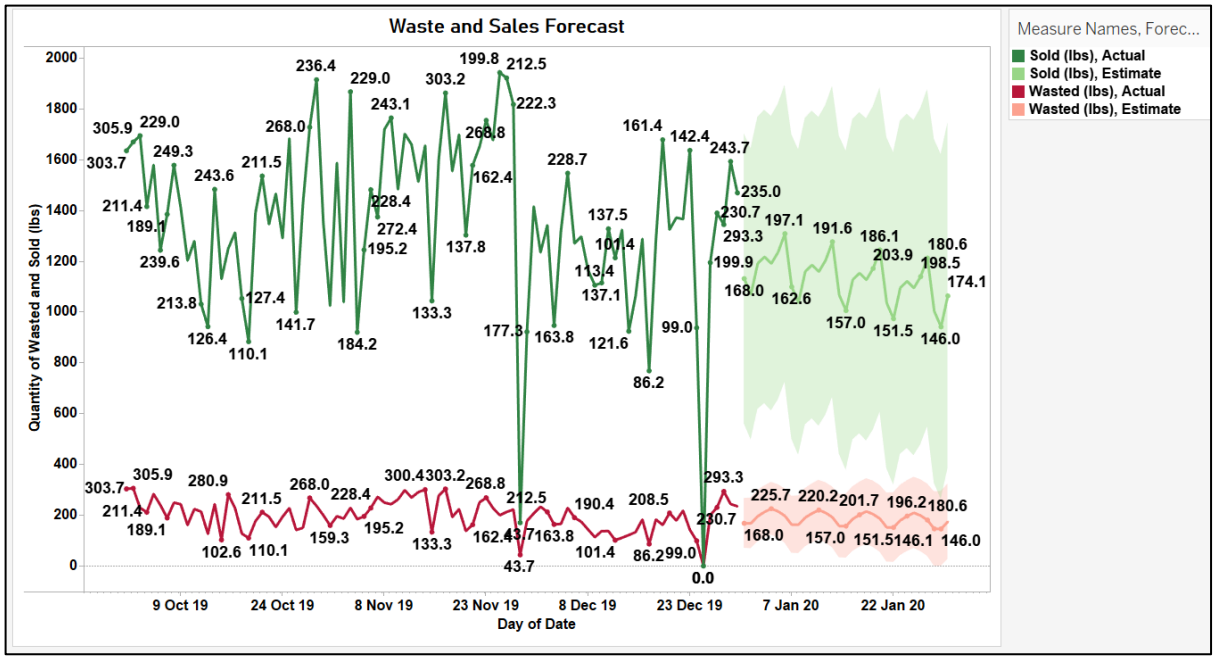


Figure 23: Dashboard2- Waste and Sales Forecast

6. The fifth graph is a clustering graph that helps the end-users analyze which food items are bad performing items, good performing items, or average performing items.

   For example, in the below screen we can see that the green cluster represents the good performing items whereas the red cluster represents the bad performing items and those are the items that need to be improved, in other words, those are the items that the user can understand are bad, and the production of those items can be reduced accordingly.

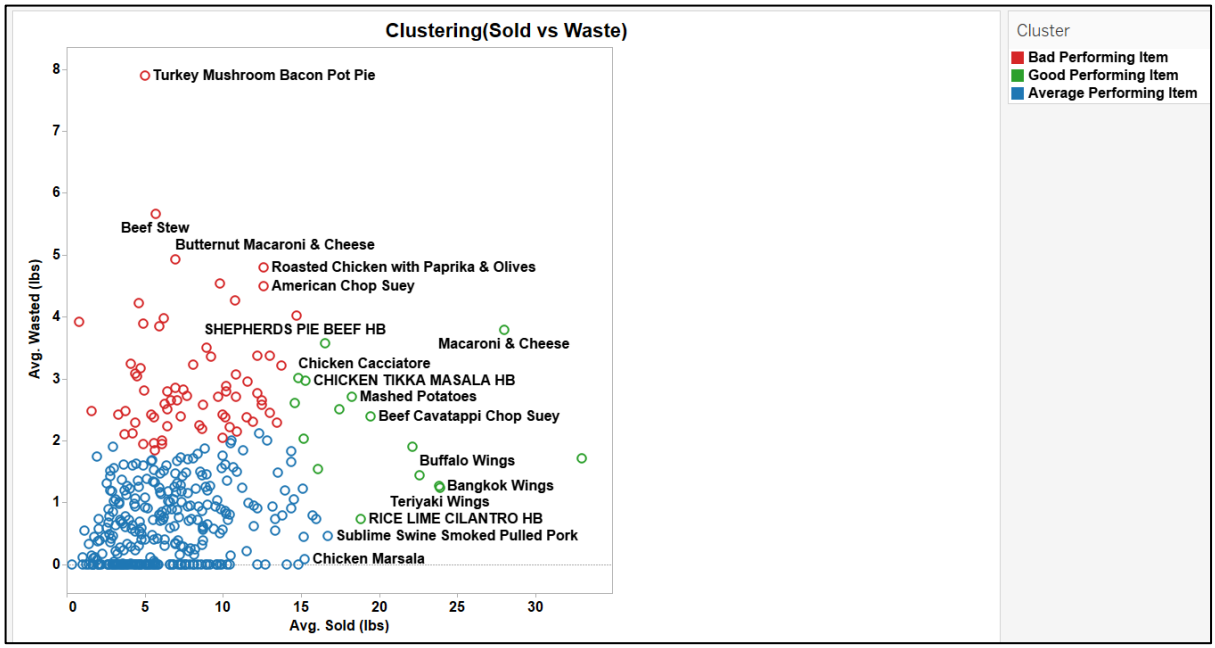


Figure 24: Dashboard2 - Clustering (Sold vs Waste)

7. For making the Dashboard2 more interactive and user-friendly, we have added several interactive and visualization features and effects.

    - We have added an action filter on the different sheets of the dashboard, which helps the end-users to see the correlation and visualize data more intensively.

    - We have also added a Reset, Navigate, and Theme Button. The reset button helps refresh the dashboard to normal if any filter was applied to it earlier by the user. The Navigate button helps the users to navigate to the second dashboard. The Theme button has been developed as an optional feature to give users the benefit of viewing the dashboard in dark mode.

    - Below are the screens that display the features and effects added to Dasboard2:

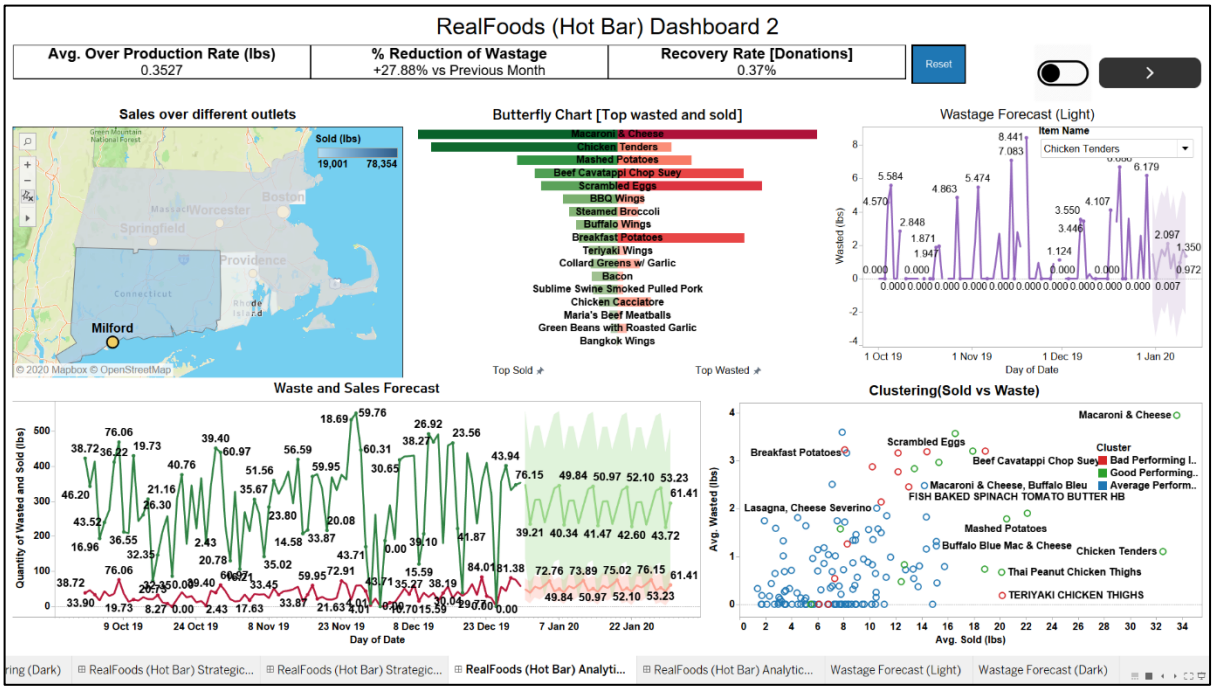


Figure 25: Dashboard2 Light Theme

Here, in the example above, we have selected Milford, CT, in the graph and can see that the action filter enables the user to see data only related to Milford in other graphs as well.

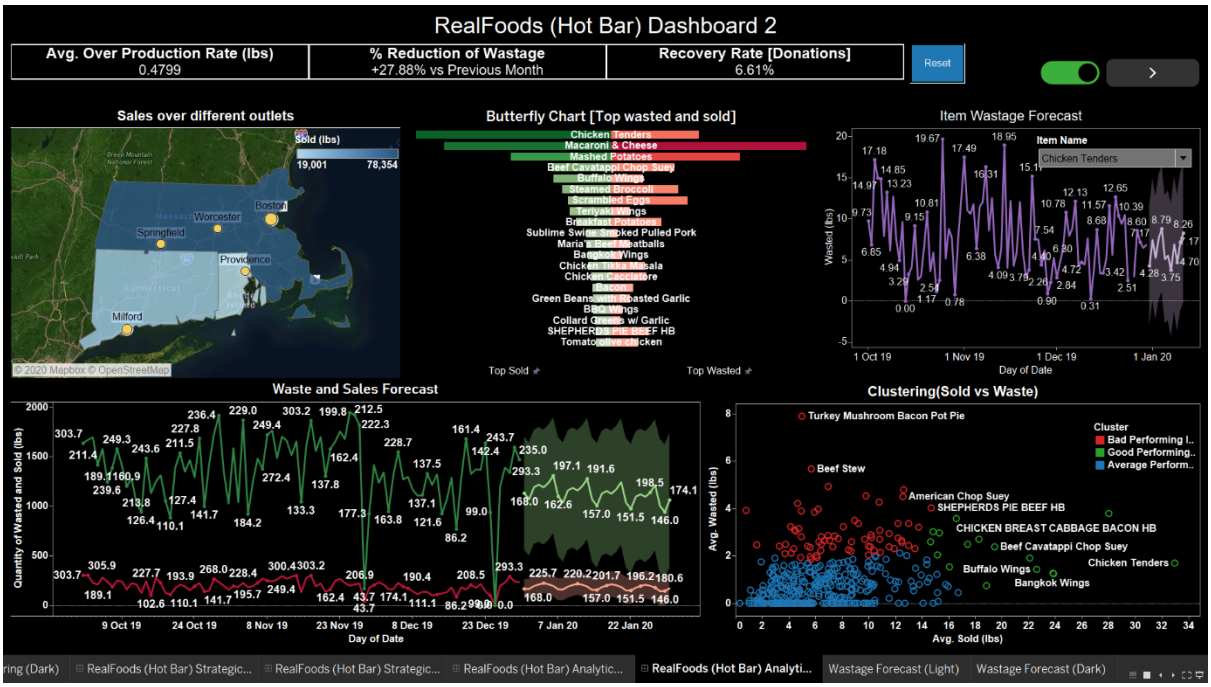


Figure 26: Dashboard2 Dark Theme

Here, in the graph above, we have selected the Dark Mode of the Dashboard2, which is an optional feature for the end-users.

Dashboard2 will help store and regional managers answer the below managerial questions:

- Which outlet/location has the most sales?
- Which are the top-selling and wasted items?
- What is next month's forecast for wastage for a particular item?
- What is next month's forecast for sales and wastage?
- Which items are good performing, bad performing, and average performing items?

4.3 Hot Bar Statistical modeling

Our **third proposed analytic approach** is a time series forecast model for predicting future trends for the top 10 items wasted using Python. For this, we used the same data set as used for our earlier dashboards. However, we cleaned the data and pre-processed it accordingly. We performed two models for predicting the forecast of the top 10 wasted food items for the next 15 days.

The first model used is **SARIMA**. Seasonal Autoregressive Integrated Moving Average, SARIMA, or Seasonal ARIMA, is an extension of ARIMA that explicitly supports univariate time series data with a seasonal component. It adds three new hyperparameters to specify the autoregression (AR), differencing (I), and moving average (MA) for the seasonal component of the series, as well as an additional parameter for the period of the seasonality.

The second model used is the **Prophet**. The Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well.

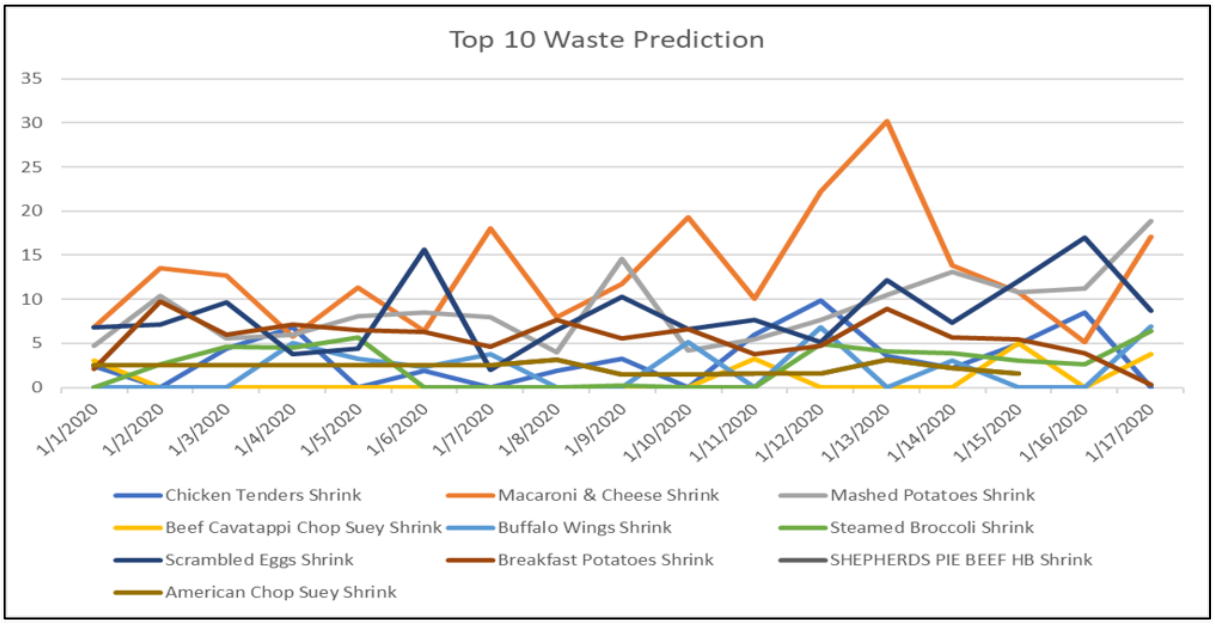


Figure 27: Top 10 Waste Prediction Line Chart

5. Implementation

Without an excellent implementation strategy, the BI solution, which including the data warehouse we developed, will be going to drown because there are all sorts of problems that can be a holdback from top management to employees at the ground level. For instance, senior management did not realize the scope of the project and underestimated the consequence, or employees unwilling to change, on account of too comfortable on the system they are using for decades. On the managerial aspect, Kotter listed a process that can help business organizations to prevent failure during transformation, and we high recommendations Real Foods follow the guideline of Kotter's eight steps model throughout the program.

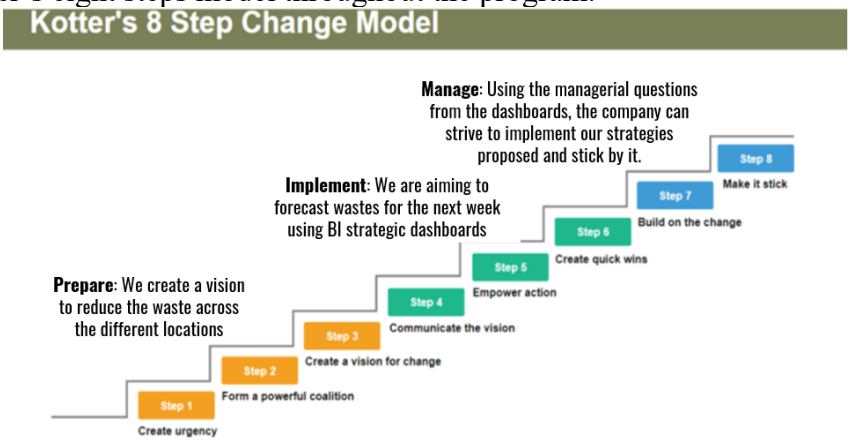


Figure 28: Kotter's Eight Steps for Real Foods

5.1 Kotter's Eight Steps

According to the framework of *Kotter's eight steps to transforming your organization* (Watson, Goodhue&Wixom,2001, p.493). The first phase Real Food will confront is be modest, do not think the company has always been doing well, lockdown in the comfort zone without leap forward, and 'establish a sense of urgency is essential.' We should highlight the number that can influence all the stakeholders, such as a statistic shows Real Food dump around 30,000 pounds of food per year in each store, the fact may not be visible, but the number cannot lie. Frequent communication also a vital part of this phase, utilizes different channels, such as meeting, email, speeches, to inform all the stakeholders about the significant problem and give a vision or plan the company can stick on it. Besides talking about the benefits on the company level, we should come up with a way that gains participants' interests and create an energetic work environment to fulfill the upcoming changes.

While everyone in Real Foods knows the urgency, 'form a powerful guiding coalition' to be able to push the project forward and under control. Bring relevant people together in the team, contains not only top management but also the management on the middle, supervisor, and expert on the technical side to commit to handling the BI project well. And as the project itself, Real Food should determent the level of importance and arrange the company's resources to support the project in terms of talents, budget, equipment, facilities, or anything that can help the transformation process.

What is the company's vision? What needs to be changed? How can we realize the vision? Finding answers to these questions will contribute to the furtherance of Real Food more likely to achieve future success. According to *The benefits of data warehousing: why some organizations realize exceptional payoffs*, "… lacking a plausible way out of a bad situation, individuals and organizations have a great capacity for ignoring disturbing evidence and continuing to act in established patterns" (Watson, Goodhue&Wixom,2001, p.494). Furthermore, we suggest Real Food 'establish a change vision,' on the topic of reducing the waste across different locations. For the purpose of management, the vision can tie with stakeholders' interests, and also the well-known knowledge amount of all employees in the company.

The knowledge behind the vision is more indispensable 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 taking related to everyone in the Real Foods.

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 the 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 implement the system will help business growth in the long-run.

On the 'plan for and create short-term wins' phase, Real Food should achieve some benefits from implementing the BI system. Such as the percentage decrease in wastage every week, and the percentage increase in donations every week. The result will be the best cure 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. On 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 Real Foods. However, we can still find out drawbacks to the system. The data we have collected may not be precise, create new KPIs fits the company goal, new visualization model that can support decision making in the Real Food.

The final phase in *Kotter's eight steps*, 'Institutionalize the new approaches'. The best result of the BI solution is part of the organization strategy, this strategy can change over time, while the BI solution continues to iterate and meet the company's needs.

5.2 Data Quality

From the technical side, data quality refers both to the characteristics associated with high-quality data and to the processes used to measure or improve the quality of data (Henderson, Earley, Sebastian-Coleman, Sykora, & Smith, 2017). And "garbage in, garbage out (GIGO)," it stands bad input will result in a bad output, and it is always a concept every company needs to follow to ensure the data quality.

Data quality is the core before we start to analyze. Like we did in this project, we need to improve the quality of data by clean the dataset, we removed all the garbage data contains null values, negative values, and other unstructured values, and refined those columns that we need on the Hot Bar section. Due to data integrity issues in January, we selected data ranges from October to December. We did the cleaning on excel, but as long as we develop Data Warehouse in Real Food, all relevant data will collect by ETL tools, and also clean and re-structure on SQL basis, since we use large dataset not only three months.

Nowadays, most data analysts or data scientists spend a significant amount of time doing data cleaning. To minimize time and cost, data governance can be a solution to ensure high-quality data. First of all, we should control the sources where data received from. Such as the kitchen in Real Foods, the food scale needs to be an electronic device that can automatically connect to our data warehouse without manual work. The process can ensure data integrity and avoid unstructured data. Secondly, set up regulation to standardize the data format that can be better using on the current BI system and fulfill our goal. Last but not least, Real Foods should hire talents and form a data management team or partner with a third-party expert to improve data quality. At the same time, use the BI solution to help Real Foods make decisions align with the company vision.

6. Conclusion

In general, without business intelligence solutions, Real Foods wouldn't able to understand which food is top sold and top wasted and what not. And while Hot Bar section of Real Foods is on teenager stage in terms of BI maturity model, adopting BI solutions can improve the overall quality of hot bar in long run such as able to learn about future based on predictive analysis and forecasting which later can set their next target in terms of advertising certain products and targeted offers according to our dashboards for each of customer's choices. In addition to it, sales revenue can be increased a lot where BI will help to know which item is most profitable and which one isn't, eliminating higher costs over time while increasing profits. BI's major components are OLAP, CPM, RTBI, Data Warehousing, and Data Source in which we proposed three things for Hot Bar section of Real Foods which was External and Operational

Data, Extraction, Transformation, and Loading, Data Warehouse and Data Marts, Monitoring and Alerting, Data Analytics, and Data Visualization. Most of these proposed BI components will help managers in Hot Bar section of Real Foods which allows them to manage their food waste as well as better decisions for preventing it in the first place. We have proposed three dashboards in the prototype stage where we feel it will help them to know which location has most sales, which food is top selling and top wasted, waste limit, and so forth. Finally, even though BI solution can be implemented, there are some risks as it all depends on Real Food's management as some officials in top management may not realize the scope of the project and some may underestimate it due to sticking to their "past" as in staying in the comfort zone of their system that one using for years or decades.

So in conclusion, we recommend Real Foods to analyze themselves as a whole in terms of staffing and management and bring relevant people together to ensure a smooth transition to next step during their transformation, or rather, we strongly recommend the company to read the framework of Kotter's eight steps to transforming your organization. By reading the framework and giving a thorough evaluation of the company as a whole, the implementation of BI solutions can become easier without any issues. Once all of the problems are solved with no future issues, we later can recommend trying out our prototype dashboards and discuss with other management to see which dashboard will help Hot Bar section of Real Foods in a long stable run before the implementation of our proposed BI solution.

7. References

Food Waste Ends Here. (n.d.). Retrieved from

https://www.phoodsolutions.com/

Food Waste FAQs. (n.d.). Retrieved from

https://www.usda.gov/foodwaste/faqs

Henderson, D., Earley, S., Sebastian-Coleman, L., Sykora, E., & Smith, E. (2017). *DAMA-DMBOK: Data management body of knowledge*. Basking Ridge, NJ: Technics publications.

Leanpath Food Waste Prevention Solution. (n.d.). Retrieved from

https://www.leanpath.com/

Pratt, M., & Fruhlinger, J. (2019, October 16). What is business intelligence? Turning data into business insights. Retrieved from

https://www.cio.com/article/2439504/business-intelligence-definition-and-solutions.html

Understanding the Major Components of Business Intelligence. (2019, August 30). Retrieved from https://www.softwareworld.co/understanding-the-major-components-of-business-intelligence/

Watson, H. J., Goodhue, D. L., & Wixom, B. H. (2002). The benefits of data warehousing: Why some organizations realize exceptional payoffs. In *Information & management*. Elsevier.

Whole Foods Market. (2020, May 02). Retrieved from

https://en.wikipedia.org/wiki/Whole_Foods_Market

Yvanovich, R. (2018, July 30). How Whole Foods Reinvents Itself with a Cloud Retail Suite. Retrieved from https://blog.trginternational.com/beyond-the-amazon-deal-how-whole-foods-reinvents-itself-with-a-cloud-retail-suite

Appendix

```
/* Code for Statistical Modelling using SARIMA and Prophet*/
import pandas as pd
data = pd.read_excel('BIdata.xlsx')
data.head()
type(data.LoggedTime)
data['LoggedTime'] = pd.to_datetime(data['LoggedTime'])
data['Date'] = data['LoggedTime'].dt.strftime('%Y-%m-%d')
data.head()
data1 =
data.groupby(['LocationId','Station','Date','ClientID','ItemName','Location'],as_index=False).sum()
data1 = data1.drop(columns=['Unnamed:
0','ID','Quantity','PanID','PanWeight','EmployeeName'])
data1
a = data1[data1['Station'].str.match('H')]
a = a[(a['Served']>=0) & (a['Saved']>=0) & (a['Sold']>=0) & (a['Shrink']>=0)]
a['Station'].unique()
a['Station']='HB'
a['Station'].unique()
x = a[a['ClientID']==46000022020]
x=x.sort_values(by=['Date'])
x= x.groupby('Date').sum().reset_index()
x=x.drop(columns=['ClientID','LocationId'])
x['Date'] = pd.to_datetime(x['Date'], format='%Y-%m-%d')
```

```
z =x[['Date','Shrink']]
z['Date'] =  pd.to_datetime(z['Date'], format='%m/%d/%Y')
z = z.sort_values(by=['Date'], ascending=[True])
z.set_index('Date', inplace=True)
z = z.resample('D').ffill().reset_index()from pandas import Series
from matplotlib import pyplot
pyplot.figure(1)
pyplot.subplot(211)
y.Shrink.hist()
pyplot.subplot(212)
y.Shrink.plot(kind='kde')
pyplot.show()
pd.plotting.register_matplotlib_converters()
rcParams['figure.figsize'] = 18, 8
decomposition = sm.tsa.seasonal_decompose(y, model='additive',freq=12, extrapolate_trend =
12)
fig = decomposition.plot()
plt.show()
ts = y.Shrink - y.Shrink.shift()
ts.dropna(inplace=True)
pyplot.figure()
pyplot.subplot(211)
plot_acf(ts, ax=pyplot.gca(),lags=30)
pyplot.subplot(212)
plot_pacf(ts, ax=pyplot.gca(),lags=30)
pyplot.show()
train = y[:int(0.86*(len(y)))]
valid = y[int(0.86*(len(y))):]

#plotting the data
train['Shrink'].plot()
valid['Shrink'].plot()

from statsmodels.tsa.arima_model import ARIMA
from sklearn.metrics import mean_squared_error
from math import sqrt

# fit model
model = ARIMA(train, order=(1, 1, 1))
model_fit = model.fit(disp=1)

model_fit.summary()

start_index = valid.index.min()
end_index = valid.index.max()
```

```
#Predictions
predictions = model_fit.predict(start=start_index, end=end_index)


# report performance
mse = mean_squared_error(y[start_index:end_index], predictions)
rmse = sqrt(mse)
print('RMSE: {}, MSE:{}'.format(rmse,mse))
plt.plot(y.Shrink)
plt.plot(predictions, color='red')
plt.title('RMSE: %.4f'% rmse)
plt.show()
predictions_ARIMA_diff = pd.Series(predictions, copy=True)
print (predictions_ARIMA_diff.head())
predictions_ARIMA_diff_cumsum = predictions_ARIMA_diff.cumsum()
print (predictions_ARIMA_diff_cumsum.head())
predictions_ARIMA_log = pd.Series(valid.Shrink.iloc[0], index=valid.index)
predictions_ARIMA_log =
predictions_ARIMA_log.add(predictions_ARIMA_diff_cumsum,fill_value=0)
predictions_ARIMA_log.head()


import numpy as np
plt.plot(y.Shrink)
plt.plot(predictions_ARIMA_log)
plt.title('RMSE: %.4f'% np.sqrt(np.nansum((predictions_ARIMA_log-ts)**2)/len(ts)))


from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error,
median_absolute_error, mean_squared_log_error
def evaluate_forecast(y,pred):
    results = pd.DataFrame({'r2_score':r2_score(y, pred),
                            }, index=[0])
    results['mean_absolute_error'] = mean_absolute_error(y, pred)
    results['median_absolute_error'] = median_absolute_error(y, pred)
    results['mse'] = mean_squared_error(y, pred)
    results['msle'] = mean_squared_log_error(y, pred)
    results['mape'] = mean_absolute_percentage_error(y, pred)
    results['rmse'] = np.sqrt(results['mse'])
    return results
def mean_absolute_percentage_error(y_true, y_pred):
    return np.mean(np.abs((y_true - y_pred) / y_true)) * 100


evaluate_forecast(y[start_index:end_index], predictions_ARIMA_log)
```

```
#building the model
from pmdarima import auto_arima
model = auto_arima(train, trace=True, error_action='ignore', suppress_warnings=True)
model.fit(train)

forecast = model.predict(n_periods=len(valid))
forecast = pd.DataFrame(forecast,index = valid.index,columns=['Prediction'])

#plot the predictions for validation set
plt.plot(y.Shrink, label='Train')
#plt.plot(valid, label='Valid')
plt.plot(forecast, label='Prediction')
plt.show()

evaluate_forecast(valid, forecast)

train_prophet = pd.DataFrame()
train_prophet['ds'] = train.index
train_prophet['y'] = train.Shrink.values

from fbprophet import Prophet

#instantiate Prophet with only yearly seasonality as our data is monthly
model = Prophet( yearly_seasonality=True, seasonality_mode = 'multiplicative')
model.fit(train_prophet) #fit the model with your dataframe

# predict for five months in the furure and MS - month start is the frequency
future = model.make_future_dataframe( periods=7, freq = "D")
# now lets make the forecasts
forecast = model.predict(future)
forecast[['ds', 'yhat', 'yhat_lower', 'yhat_upper']].tail()

fig = model.plot(forecast)
#plot the predictions for validation set

plt.plot(valid, label='Valid', color = 'red', linewidth = 2)

plt.show()
ans = ans.join(valid, on='Date')
ans = ans.rename(columns={"Shrink": "American Chop Suey Shrink"})
ans = ans.drop(columns=['American Chop Suey Shrink'])
ans.to_csv('top10_waste_prediction.csv')
```