**Interactive Data Visualization Dashboard for Point-of-Sales Systems Paper**

Chen Shiqi, Tan Wei Lin Joanna, Singapore Management University

# Abstract

Traditional POS systems are used mainly for completing orders and payments from a terminal. Data collected from these systems are hence limited to sales figures. A successful F&B business should not only analyse sales figures but also other aspects of the business such as efficiency of the current operations, inventory stocking, marketing campaigns and etc.

In this project, we aim to provide an interactive dashboard of different visualizations of data collected from HoiPOS systems. The dashboard will be built using d3.js and related libraries. The project aims to improve HoiPOS’ data visualizations as well as add more visualizations of data analysis to value-add their POS system to clients. This project applies descriptive analytics and effective visualizations to gain insights not only on the sales performance but also operations and marketing campaigns.

# Introduction

HoiPOS [1] provides innovative point-of-sales (POS) systems to the Food and Beverages sector. The company offers efficient and innovative tablet POS systems for their clients while giving them competitive advantages by providing useful analytics solutions to aid business owners in making business decisions. It is important for restaurants to know how to exceed customer satisfaction in order to stimulate current growth in sales through repeat customers and to attract new customers.

The data collected from HoiPOS systems is enormously extensive; it stores and records more information compared to traditional systems that usually only take into account the sales amount and profits. As HoiPOS systems are integrated into tablets, the data collected covers more than just sales figures. It includes information such as time taken to serve a certain dish, which combination of dishes are often ordered together and more.

# Motivation

Traditional POS systems are used mainly for completing orders and payments from a terminal. Data collected from these systems are hence limited to sales figures. A successful F&B business should not only analyse sales figures but also other aspects of the business such as efficiency of the current operations, inventory stocking, marketing campaigns and etc.

As HoiPOS systems are programmed to offer a broad array of features such as sending/receiving orders to/from the kitchen, modifying orders, completing orders, and completing different modes of payments for an order. The data that is collected from these functions are more comprehensive compared to data from traditional POS systems. By analyzing and making sense out of these data, it could help businesses identify trend and find out information that is unknown to them and thus helping them improve, profit and efficiency. While the current analysis is helpful as it compares historic data and real-time data to compare trends and identify pitfalls and successes in the business, the visualizations used can be improved to offer a more comprehensive understanding of the data.

Most of HoiPOS competitors do not offer visualizations of data that is actually useful to businesses. Merely identifying that a business has performed better last week than this week does not provide useful information to the owners. Knowing the cause of last week’s better performance will be more useful as they can replicate the cause to achieve success again. An example would be identifying which particular item’s popularity is contributing to the better sales performance. This allows the owner to inform the kitchen to prepare a certain dish at a faster rate which will translate to higher table turnover rate. Furthermore, most of HoiPOS competitors also do not offer comprehensive visualizations of the data analysis which often confuses the client and hence not maximizing the potential of the information extracted from the data.

# Objectives

The objective of this project is to provide an interactive dashboard of different visualizations of data collected from HoiPOS systems. The project aims to improve HoiPOS’ data visualizations as well as add more visualizations of data analysis to value-add their POS system to clients. This project also aims to be able to apply descriptive analytics and effective visualizations to gain insights not only on the sales performance but also operations and marketing campaigns.

The final product will consists of an interactive visualization dashboard which will be designed to provide visualizations that can be easily understood by the layman as well. Intuitive functions such as cross filtering, will be implemented to allow business owners to conduct exploratory data analysis without needing technical knowledge. To ensure that the dashboard is easy to use, the team will also conduct several user testing with stakeholders.

The main objective of the project would be to develop the following functions/visualizations:

### Filter Data by Date/Time

* This allows the user to choose certain periods of the year to analyse as different seasons of the year could have a higher anticipated sales performance e.g restaurants near F1 location during F1 event
* Being able to view analysis of data at a more micro level also allows the business owner to understand which time periods are peak periods and hence make better human resource deployment decisions quickly
* Anomalies in the data can also be noticed quickly and will be able to have a reference point of why the anomaly happened using the date and time. E.g Spike in orders during F1 event

### Cross-Filter

* Allow users to quickly have a glance of different set variables across a few different visualizations.

### Dashboard Functionalities

* Sales Performance over time:
  + Includes overview of sales, amount earned a day, sales performances, target analysis and more
  + Compare historic data and real-time data to notice gaps and/or trends
* Association Analysis of items
  + To allow users to identify popular combination of items. This information is useful when crafting marketing campaigns e.g combo meals of the popular combinations of items
* Productivity Analysis
  + Users will be able to analyse time taken to prepare different food items and how it varies across the day. This insight will be able to aid business owners in making operational decisions with regards to human resource and productivity in the kitchen.
* Popular Items and Modifiers Analysis
  + To allow users to identify most popular items and/or identify items that are least popular. This information informs users about items that are not being ordered frequently could be removed from the menu to reduce wastage and costs
  + Exploring further for Popular Items - Visualization of Popular Items Modifier
    - To allow users to identify the most popular item modifier i.e hot/cold options for a cup of tea. This information may prompt users to modify the menu to accommodate for the popular preferences.
    - Users will be able to cross-filter the data using time as a variable to see the popular items at different time periods

# Technologies used

* D3.js [2] - D3.js is a JavaScript library for manipulating documents based on data.
* DC.js [3] - dc.js is a javascript charting library with native crossfilter support and allowing highly efficient exploration on large multi-dimensional dataset (inspired by crossfilter's demo).
* JMP Pro [4] - JMP Pro is a statistical analysis software from SAS that provides all superior capabilities for interactive data visualizations, exploration, analysis and communication.
* Tableau [5] - Tableau Software helps people see and understand data. Tableau helps anyone quickly analyze, visualize and share information.
* Microsoft Excel [6] – A spreadsheet software developed by Microsoft that allows users to perform data manipulation.

# mETHODOLOGY

For this project, the methods used in data analytics are descriptive analytics through data visualization. Data visualization is the presentation of data in a graphical or visual format where it allow readers to visually interpret analytical results easily and grasp difficult concepts or identify new patterns. [7] Data visualizations represents data by taking advantage of the unique ability of visual perception to detect meaningful patterns that might otherwise be hidden.

Furthermore, with interactivity, it allows users to further drill down into charts and graphs for more details, changing the data being interpreted.

The data visualizations that are included in our dashboard are: Line charts, bar charts, row charts, pie charts, area charts, sunburst charts and heat map chart. The interactivity aspect of the project utilizes cross-filter to allow users to filter data across charts depending on their needs e.g monthly vs yearly sales performance.

# dATA cLEANING

## Item Table

Plu are supposed to be unique but duplicate plu was found in the item table. The team clarify this with the sponsor and understand that this happen because the store side added in the plu without checking that the plu already exist in the system. As there are only 8 of such instance, the sponsor said that the team could manually edit the plu to some other value.

## ​Order Table

In the Order table, we discovered only 1 outlier in the created\_timestamp column where the data is from the year 2019. Since this is obviously a data entry error, we removed it from our table. There were 2 other outliers from the food\_preparation\_duration column where the time taken to prepare the food took more than 600 minutes. In the business context of our dataset, that is not possible. Hence, these rows are removed from the table.

In the total column, there were 2 outliers with the values 500. To verify if this is valid, the team cross-referenced the data to order\_item\_parent table using the order\_id variable. The data rows are missing from the latter table and hence, deeming these data rows to be invalid. These rows could be created as testing data.

## ​Order\_item\_parent Table

In the order\_item\_parent table, there were 2 outliers from the quantity column with values 15 and 20. Upon further investigation by checking the item\_name, the team decided that this is relevant and should be retained in the data due to the nature of the item that is purchased.

There were 1465 missing PLU codes in this table. After discussing with the sponsor, the team understands that some of the data rows with missing PLU codes are open items that are created on an ad-hoc basis. To verify if all of these data were open items, the team cross-referenced these data rows with their item\_id and item\_name. Open items should not have item\_id as well. However, the team found that only 125 of these selected data are open items. The other data rows with item\_ids are cross-referenced and updated with the other present data in the table by matching their item\_ids.

## ​Order\_item\_parent\_option Table

In the last table, the team removed 5 rows with null item\_id as upon discussion with the sponsor, the team decided that these rows are data entry errors and are anomalies. There were also 213 missing PLU values. Again, to verify whether these data rows are still relevant, the team cross-referenced the other variables. We find that the item\_ids are not missing and hence update theses rows’ PLU codes by matching the item\_ids with existing data rows in the table. After cleaning, there are no missing values for PLU column.

# dATA mANIPULATION

## Sales Performance Analysis

The final data use for Sala Analysis consist of the following fields:

|  |  |  |
| --- | --- | --- |
| Variable | dATA FORMAT | vARIABLE tYPE |
| date | Numeric | Continuous |
| close | Numeric | Continuous |

Table 1 Data description of data used for Sales Performance Analysis

The date will be formatted using Javascript to extract the month and day for use in the dashboard. Close refers to the total sales amount generated for the day and Excel pivot table is used to aggregate this value.

## Association Analysis

The final data we use for Association Analysis consists of the following fields:

|  |  |  |
| --- | --- | --- |
| Variable | dATA FORMAT | vARIABLE tYPE |
| categories | String | Nominal |
| size | Numeric | Continuous |
| date | Numeric | Continuous |

Table 2 Data description of data used for Association Analysis

The categories variable refers to the combination of categories ordered e.g 31-32-end and size refers to the number of orders containing this set of combination. Using JMP Pro, we first perform a simple distribution analysis of order\_id as the y-axis and grouping them by the category\_id. This shows us the category\_id for each order\_id. Using JMP Pro, we can right click the results of the distribution and convert it into the data table. We then replace the count with the category\_id itself as we are only interested if the category\_id is present or not. Next, we remove all rows with only a single category\_id as there is no association analysis for those orders.

This will give us the following outcome:

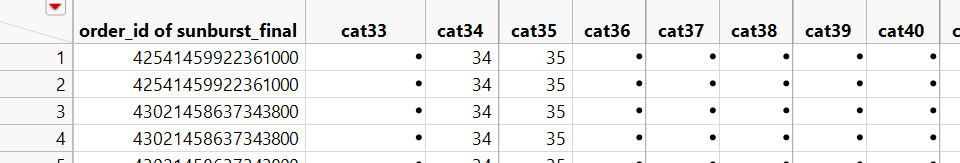


Figure 1 Screenshot of data for Association Analysis

We then use Microsoft Excel to perform a =CONCATENATE() of all category\_id columns which produces the following final data:

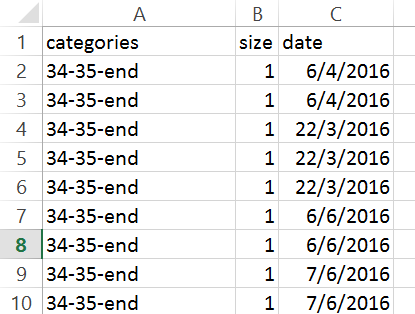


Figure 2 Screenshot of final data used for Association Analysis

Each categories variable has an additional “-end” due to a limitation of the library. Without the –end, the library is not able to create hierarchies of items with multiple children nodes.

## Productivity Analysis

The final data use for Productivity Analysis consist of the following fields:

|  |  |  |
| --- | --- | --- |
| Variable | dATA FORMAT | vARIABLE tYPE |
| categoryid | String | Nominal |
| formattedDate | String | Nominal |
| fulldate | Numeric | Continuous |
| PrepTime | Numeric | Continuous |
| Hour | Numeric | Continuous |
| Time | String | Nominal |

Table 3 Data description of data used for Productivity Analysis

Using JMP Pro, we extracted date and time from created timestamp and modified timestamp. From the extracted Time variable, we extracted the hour to a new variable named Hour. Next, we get the day of the week from the date. To display the day in an orderly way in our dashboard, we created a formattedDate variable. The formattedDate variable value example is “1.Sun”. As this is a relatively larger dataset, we created this variable instead of using the code to do it as compared to the data for sales. The PrepTime variable is obtained by subtracting the time of the modified\_timestamp with the time of the created\_timestamp. The Time variable is a formatted string to show the hour of the day in 12 hour format.

## Popularity Item and Modifier Analysis

The final data we use for Association Analysis consists of the following fields:

|  |  |  |
| --- | --- | --- |
| Variable | dATA FORMAT | vARIABLE tYPE |
| items | String | Nominal |
| size | Numeric | Continuous |
| date | Numeric | Continuous |

Table 4 Data description of data used for Popularity Item and Modifier Analysis

Similar to the Association Analysis set, the items refer to the combination of option\_plus (i.e item modifier) tied to each item\_parent and the size referring to the number of orders containing that set of combination. Unlike the association analysis data, there are too many item\_parent\_plus to do a distribution analysis. Hence, we used Microsoft Excel and the =IF() function to iterate through the data. We first sort the data according to the order of: 1. Item\_parent\_id, 2. Item\_parent\_plu and 3. Option\_plu. Next, using =IF(), we check if the next row of option\_plu belongs to the first item\_parent\_id by matching the item\_parent\_id and item\_parent\_plus. This is method is repeated comparing the 1st row and the 3rd, 4th, 5th row and so on. Once there are no more option\_plus being matched, the process stops and we have the following data:

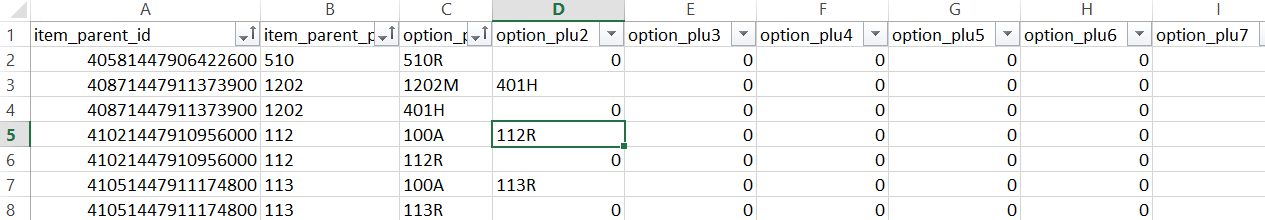


Figure 3 Screenshot of data manipulation for data used Popularity Item and Modifier Analysis

Finally, like the data for Association Analysis, we use the =CONCATENATE() function and combine all option\_plu columns together and added an “-end” to the end of each combination. This results in the following final dataset:



Figure 4 Screenshot of final data used for Popularity Item and Modifier Analysis

The first item in each items row is always an item\_parent\_plu value.

# Dashboard And Insights

## Sales Dashboard

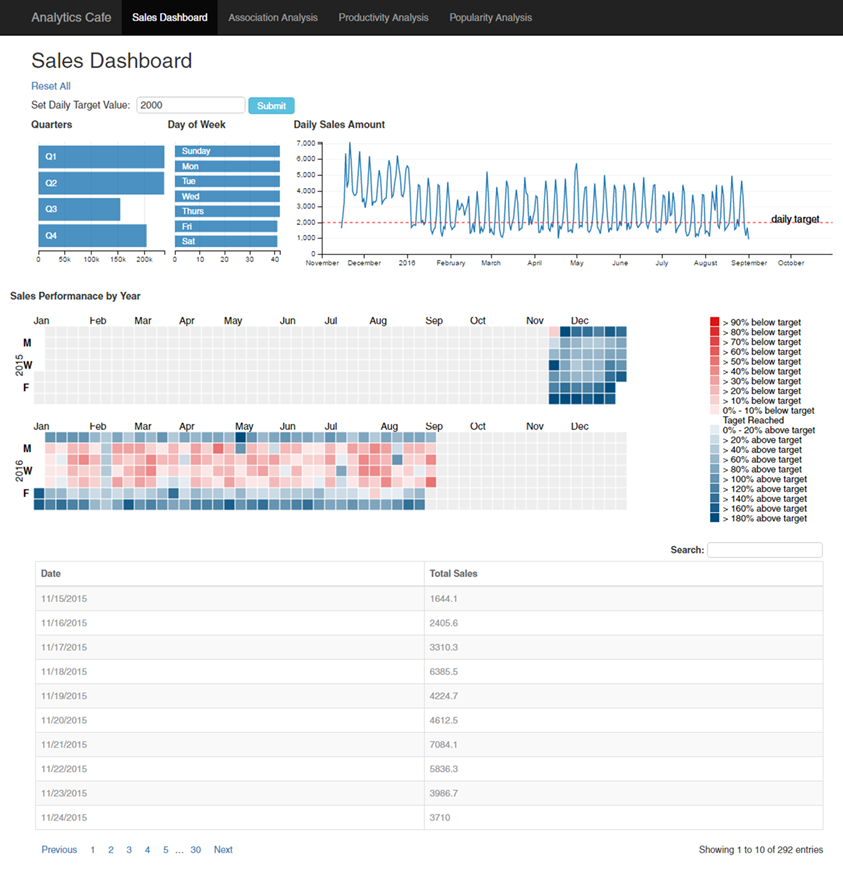


Figure 5 Sales Performance Dashboard

In the sales dashboard, there are 4 charts and 1 table to provide insights on sales data. The dashboard also has the function of cross-filter implemented to link the charts together.

The Quarters row-chart gives an overview of the total sales for each quarter of the year. From this chart alone, it shows at Q3 is the lowest performing quarter. This is not surprising as the data that we used only contains data up to September. We can also see that Q1 and Q2 have similar sales performance.

The Day of Week row-chart shows the number of days of data is present in our dataset. E.g there are 42 Sundays in the entire dataset. The main purpose of this chart is to allow the user to select a particular day to filter the data accordingly.

The Daily Sales Line chart shows the total sales collected for each day. This chart allows the user to zoom in into a specific time period to have a clearer view of the data. In this instance, we can see from the graph that the sales from November to December 2015 performs significantly better than the other months in 2016.

The Daily Sales Performance by Year Calendar Chart allows users to view the sales performance of each day of a year quickly. The color of each square represents how well/bad the sales performance is when compared to a target value. In our dashboard, we set the default target value of $2000. The user can change the target by entering the value to the textbox under set daily target value and click submit. The calendar view shade will be updated according to the target. For example, squares with shades of blue shows that the sales performance of the day is higher than $2000. The intensity shows how much higher the sales performance of the particular day was. Likewise, shades of red represents sales performance below the target.

## Association analysis

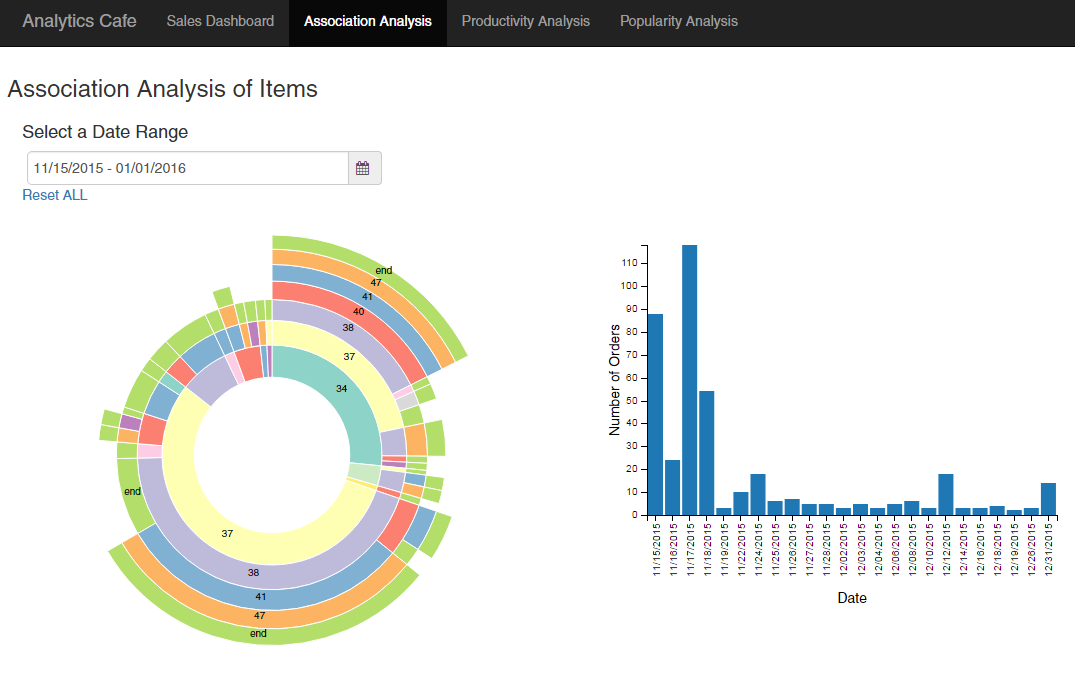


Figure 6 Association Analysis of Items Dashboard

In the Association Analysis Dashboard, users are able to explore the association of different items from the menu. The association analysis in our case is conducted between categories.

From the sunburst chart, users will be able to see which categories of items are often ordered together. For example between 15 November 2015 and 1 January 2016, items from category 37 is often ordered with items from category 38, 41 and 47. This information allows the management to create set menus that include items from the 4 categories to attract customers and hence increase sales.

From the bar chart, the users will be able to see the number of orders per day. Using the cross-filter function, users will be able to select days that he/she is interested in zooming into to get a more precise analysis of the orders ordered on the selected days.

## Producitivity Analysis

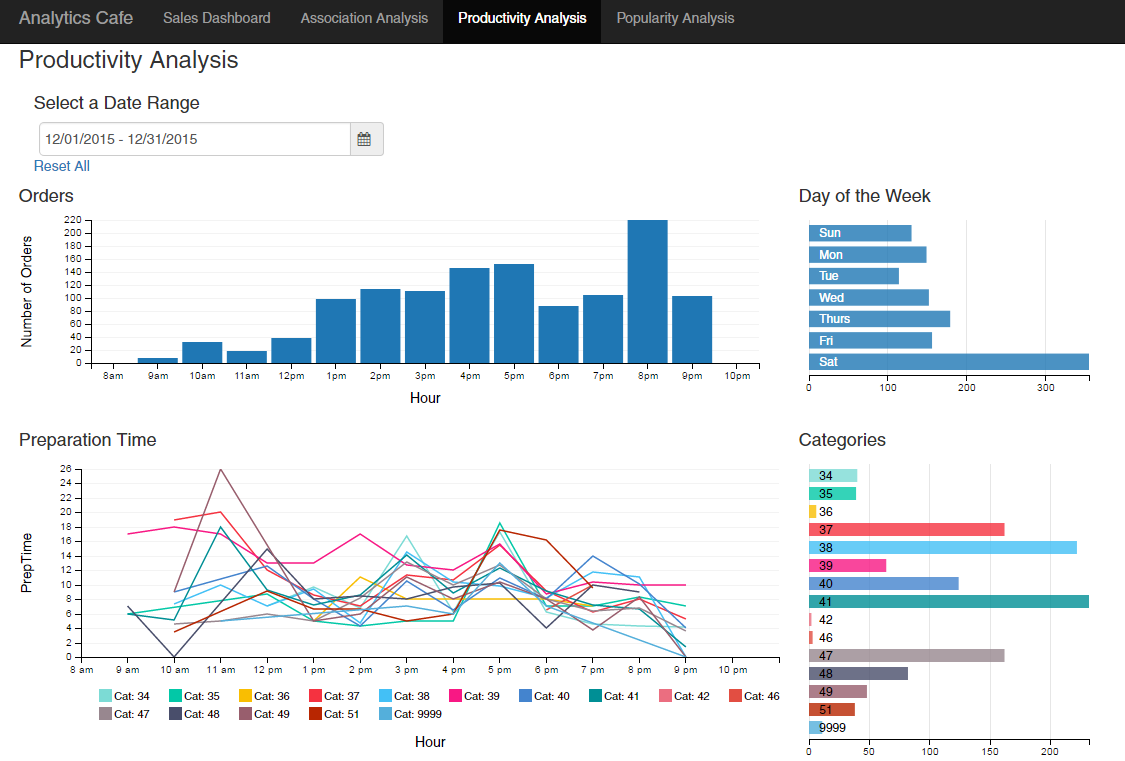


Figure 7 Productivity Analysis Dashboard

In the Productivity Analysis Dashboard, the user will be able to identify gain insight on whether the number of orders will impact the preparation time. The bar chart for Orders shows the number of order at different hour of the day. For example, in the month of December 2015, we can see that the peak period for the store is 8pm which make sense since it is dinner time.

The row chart for Day of the Week shows the accumulated number of items ordered for the specific day of the week. We can see that the day with the most number of order is Saturday followed by Thursday. Weekends usually have a higher amount of crowd but in from this chart we can see that the sales for Friday and Sunday are not performing as well. In fact, the sales for Saturday is more than double compared to the sales for the other days. This information allows the management to future investigate the reason behind this result.

The line chart for Preparation time shows the preparation time of the order for different category at different hour of the day. This can be used to identify if more orders would increase or decrease the time taken to prepare the order. The row chart for Categories is to show which category is the most popular. It is useful to help understand which categories has the longest preparation time to identify which type of resources such as equipment or manpower that should be added.

## pOPULARITY iTEM AND mODIFIER aNALYSIS

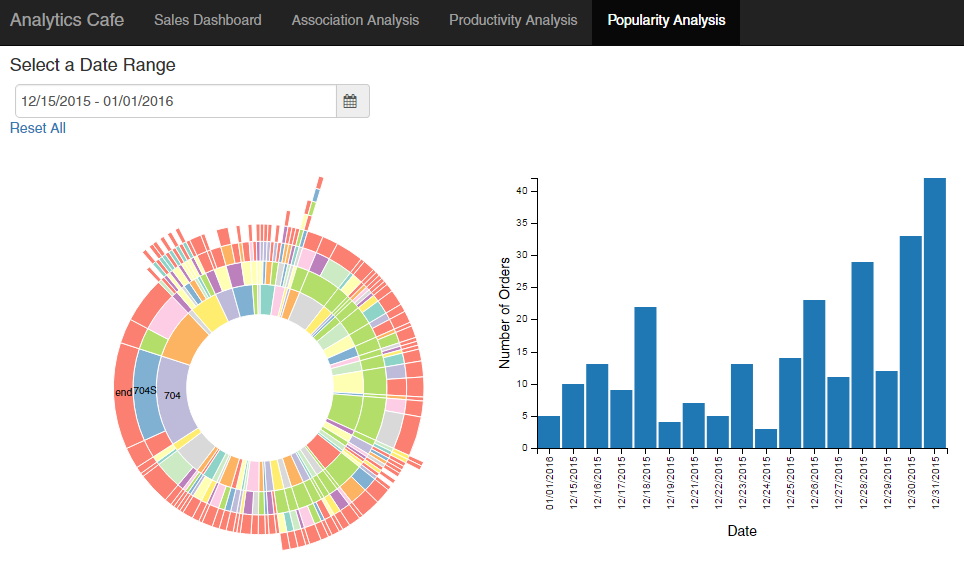


Figure 8 Popularity Item and Modifier Analysis Dashboard

In the Popularity Analysis of Items and their Modifiers Dashboard, users will be able to gain insights on the most popular types of modifiers for the item parents. The innermost arcs of the sunburst chart represents all item parents ordered for the selected date range. In this case, from 15 December to 1 January, the most popular combination for item parent 704 is the modifier 704S.

Similar to the Association Analysis Dashboard, the barchart on the right shows the number of orders per day. Users can select on the dates to zoom into the popular items and their modifiers analysis at specific dates.

## Insights from december 2015

In this section, we will be describing some of the insights from December 2015 that are gained from looking at the dashboard.

### Sales Performance

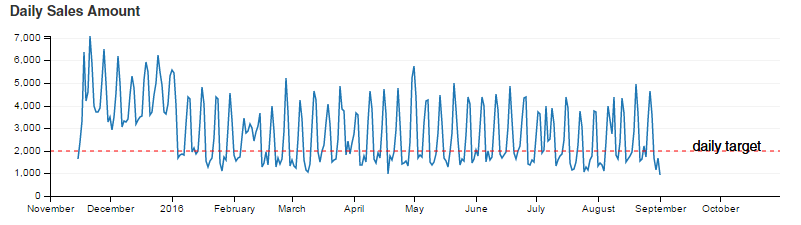


Figure 9 Daily Sales Amount

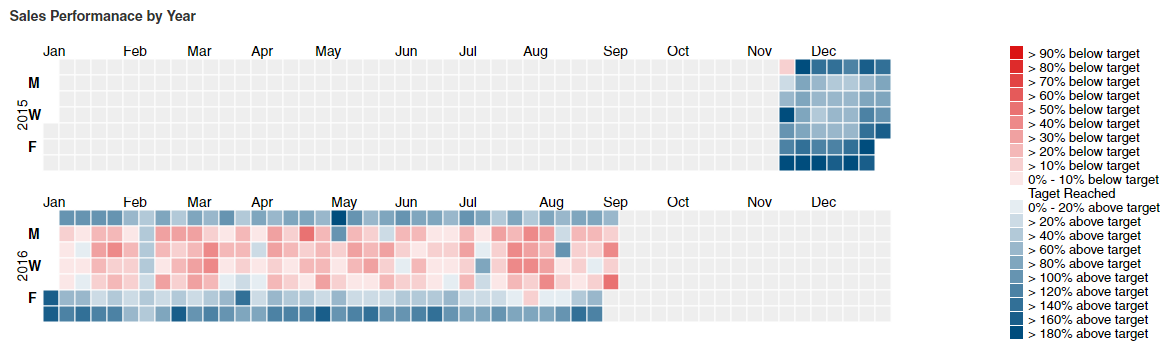


Figure 10 Sales Performance by Year in Calendar View

From the above charts we can see that December is the best performing month. From the calendar chart it shows that December is the only month where the daily sales hit the target in this scenario which is $2000 as all of the square has the blue shade. December is also the month where most of the highest daily income were recorded as there is the most number of square with darkest blue shade.

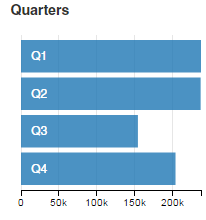


Figure 11 Total Sales by Quarters

From the bar chart we could see that the difference in the sales generated by Q4 compared to the other quarters is not very huge even when the data we have for Q4 was only from 15 November till 31 December 2015. When the dataset is full, we can definitely conclude that Q4 would be the best performing quarter.

### Top Combination of Categories

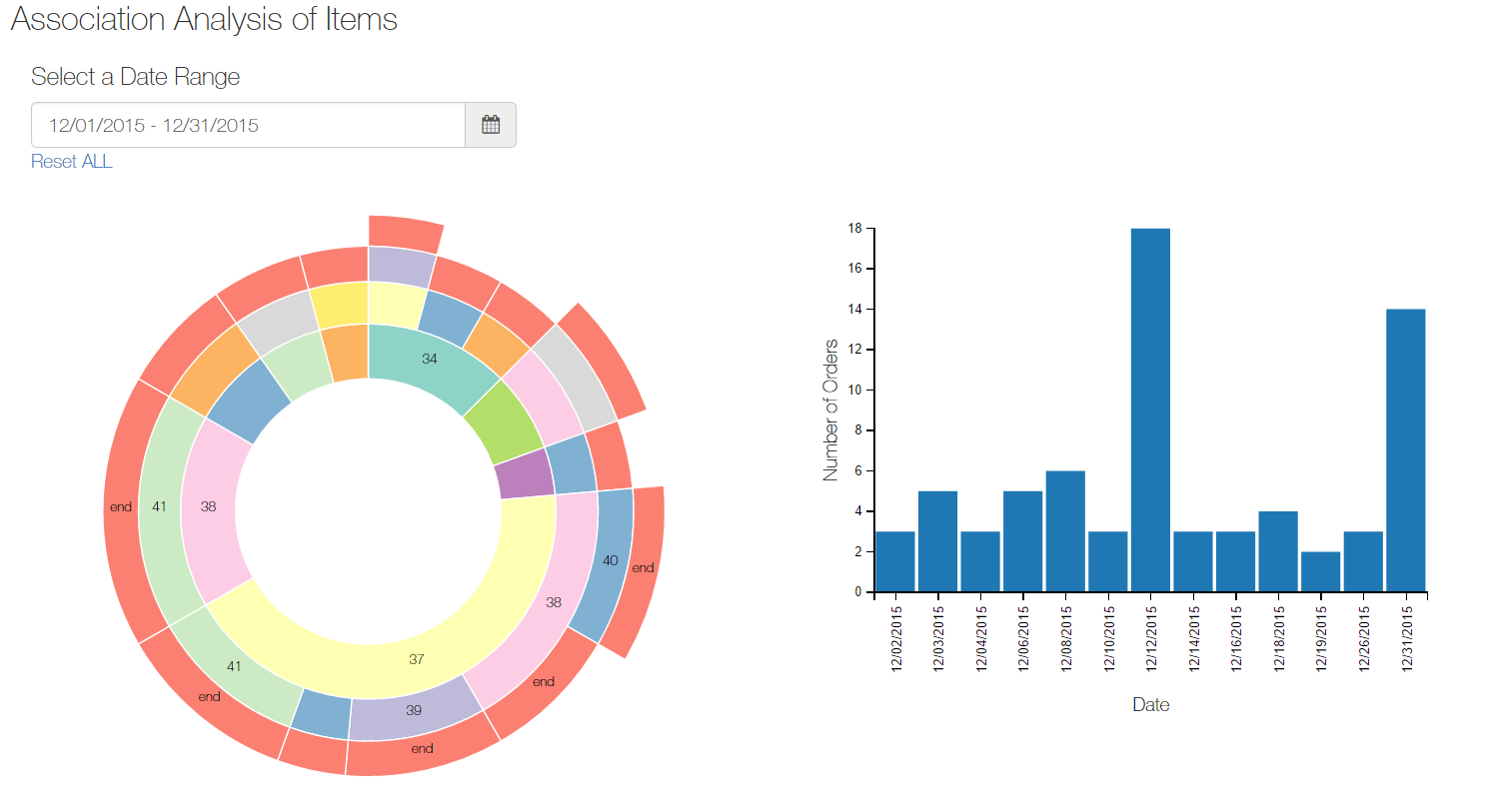


Figure 12 Association Analysis dashboard for the period of December 2015

Since December 2015 was the best performing month and the cafe’s marketing department wants to create a new promotions for the next month in hopes to increase sales. Using the date-range-picker, we select December 2015 sales data in the Association Analysis Dashboard.

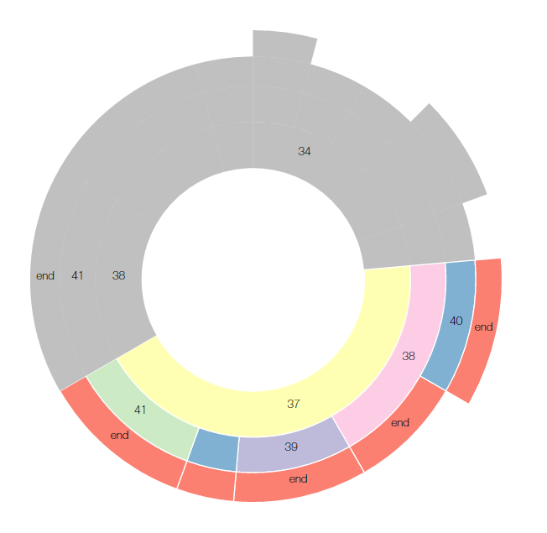


Figure 13 Association Analysis for Category 37

From the sunburst chart, we see that from the inner radius, category 37 has the largest number of orders. From there, we see what items are commonly ordered with category 37. It’s category in order to attract more customers. It seems that most customers that order items from category 37 also order items from category 38. The management can decide to create set menus for items from category 37 and category 38.

### Preparation Time During Peak vs Non-Peak Period

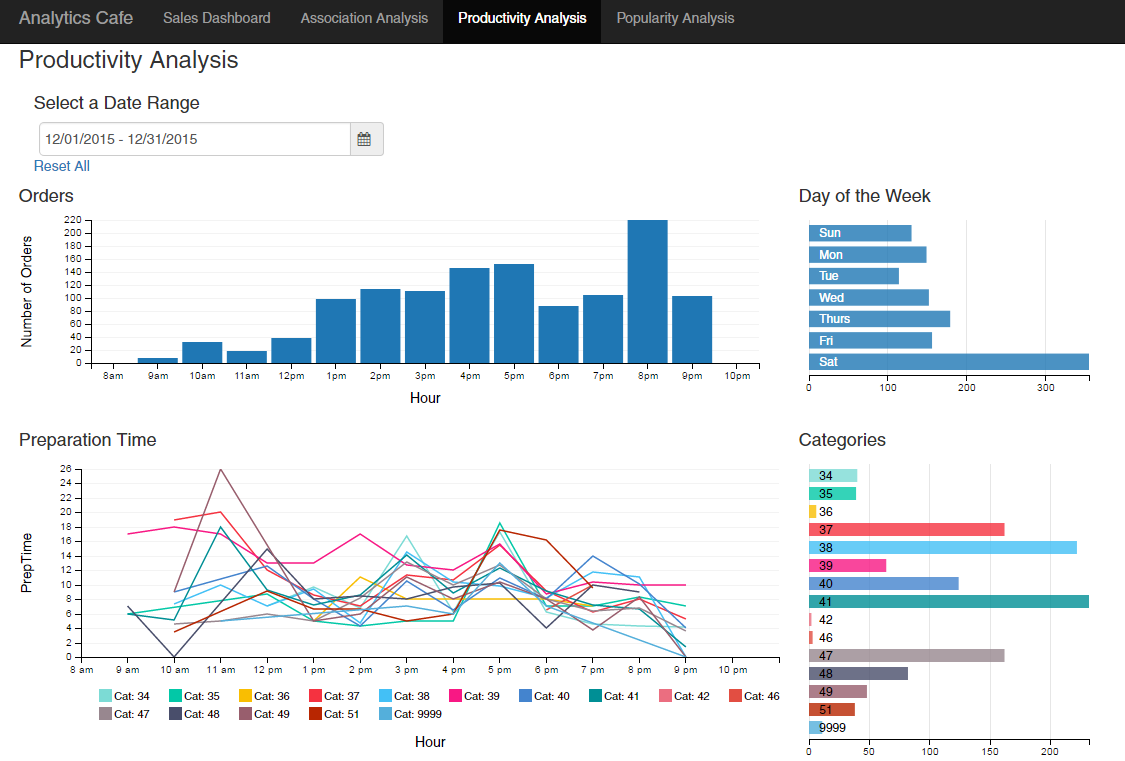


Figure 14 Productivity Analysis for December 2015

As December is the best performing month, the cafe’s operation department want to know if there are sufficient manpower to serves its customer. From the bar chart, we can see that the cafe is most crowded during dinner time and from the day of the week row chart we could see that Saturday is the busiest day of the week. More manpower will be required to ensure the productivity is not negatively affected.

The user can see that during the peak period, the preparation time decrease however, during non-peak periods, the preparation time is much higher than during the peak period. This could be an issue to look into as the kitchen might either be not performing at its most efficient during that period or there may be a lack of resources during the non-peak period.

Also, by looking at the categories row chart, the user can identify which categories are the most popular and add the resources or make adjustment accordingly.

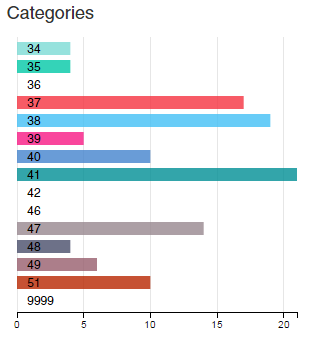
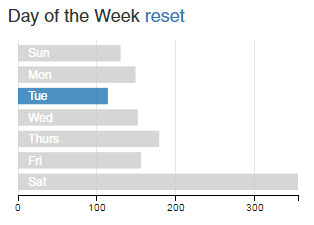


Figure 15 Category Order for Tuesday

For example, there is no or low number of order for certain category on Tuesday, the management can decide to not sell the items in that category or reduce the quantity to reduce cost and wastage.

### Top Modifier for Best-Seller Item Parent

Another way of using the dashboard for insights is that the user may identify certain item modifiers that are commonly ordered and have promotions or special discounts to increase the number of orders for that item and hence in turn increase sales performance overall.

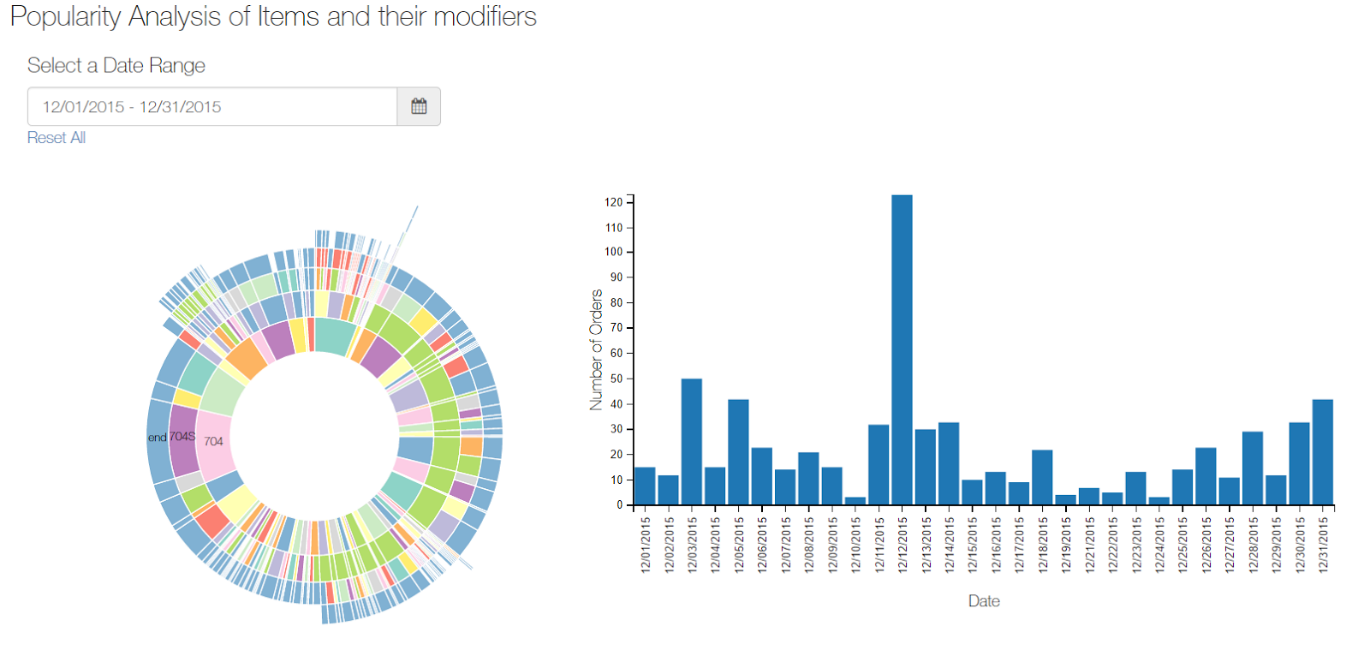


Figure 16 Popularity Item and Modifier Analysis for December 2015

After filtering the data for the month of December 2015, we can see the following chart.

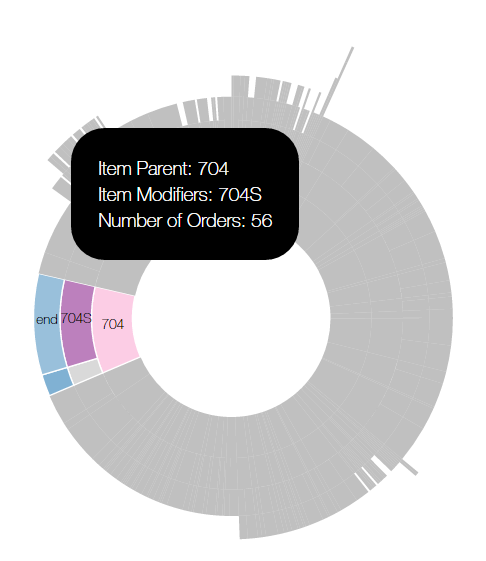


Figure 17 Popularity Item and Modifier Analysis for Item Parent Plu 704

From the sunburst chart, we can see that the most popular item parent being ordered is item 704 and the most popular item modifier option being ordered is 704s. This type of information can help the management to decide whether to promote that item with special discounts when ordered as opposed to the less popular modifier of 704D.

# Conclusion

This analytical dashboard is useful to F&B businesses as it breaks down many aspects of the business i.e sales, preparation time, popular items, and associated items. The insights gathered can definitely help the business to improve its business process if used appropriately. For example, in the productivity analysis, by intuition, preparation time should be higher during peak period due to higher volume of orders coming in. However, the data shows otherwise, showing that preparation time is actually lower during peak periods. Possible reasons to explain this event could be the kitchen do not have enough cooks during off-peak periods as well, or that the cooks do not feel a sense of urgency during these periods. Such insights are valuable as they may highlight issues that may not be expected.

While this project addresses these issues, like any other projects, there will be room for improvement as the business issues addressed in this dashboard are not exhaustive. Managers might want to have a bullet chart to represent sales performance so as to clearly identify sales performance to see how far away or ahead are they in terms of sales targets.

In the era of data analytics being one of the most highly coveted technology in businesses, HoiPOS is on the right track in satisfying client’s needs

# References

1. Ming, Q., Tay, C., Nguyen, T., & Lee, T. (2016*). HoiPOS | Truly Smart Mobile Point of Sale (POS) for Food & Beverages Restaurants, Cafes and Stores. HoiPOS.* Retrieved 25 November 2016, from http://www.hoipos.com/about-us/
2. Bostock, M. (2016). *D3.js - Data-Driven Documents*. *D3js.org*. Retrieved 25 August 2016, from <https://d3js.org/>
3. *dc.js - Dimensional Charting Javascript Library*. (2016). *Dc-js.github.io*. Retrieved 25 August 2016, from <https://dc-js.github.io/dc.js/>
4. *JMP Pro*. (2016). *Jmp.com*. Retrieved 25 August 2016, from http://www.jmp.com/en\_dk/software/jmp-pro.html
5. *Tableau* (2016). *Tableau.com*. Retrieved 25 August 2016, from <http://www.tableau.com/products/desktop>
6. *Harvey, Greg (2006).*[Excel 2007 For Dummies](https://books.google.com/?id=i5ffgL9KEg8C)*. Wiley.*[*ISBN*](https://en.wikipedia.org/wiki/International_Standard_Book_Number)[*0-470-03737-7*](https://en.wikipedia.org/wiki/Special:BookSources/0-470-03737-7)*.*
7. *Data Visualization: What it is and why it matters*. (2016). *Sas.com*. Retrieved 25 November 2016, from http://www.sas.com/en\_us/insights/big-data/data-visualization.html

# Acknowledgments

The authors wish to thank HoiPOS and Professor Kam Tin Seong.