Evaluating and Establishing KPIs and Staff Performance in F&B store
using Regression and Clustering

Peng Ho Oh, Jhun Boon Tan, Jessie Yap

Singapore Management University

Abstract

Key Performance Indicators are often used in the business world to measure an organization’s performance. However, managers often find themselves overloaded with a large number of KPIs to track and monitor. It is important therefore, to ascertain the relevance of the KPIs that an organization is tracking. In this paper, regression is used as a tool to determine the relevance of existing KPIs of Japanese F&B chain, Teppei Syokudo. We collected 720 hourly observations of different 9 different KPIs, as well as the cashiers who are present during each hour and showed that the current KPIs that the store is using is sufficient and relevant to predict Store performance.

# Evaluating and Establishing KPIs and their drivers in F&B store using Regression

Overall sales, profits, and customer volume are the basic targets for food and beverage (F&B) stores. However, the challenge comes when attempting to achieve these targets. F&B store managers are unclear as to the factors which contribute to store sales. Because of this, they are unable to focus on key factors which would boost sales. Additionally, one factor that leads to overall store performance is staff performance. The second roadblock occurs in the evaluation of staff performance in F&B stores. Which factors should staff be evaluated on, and what benchmark values should be set? Which staff are considered good performing and which bad?

## **Evaluating Factors Relating to Store Sales Using Regression**

To identify the factors (or variables) that contribute to store sales, regression analysis can be used. Regression identifies the strength and direction of relationships between dependent and independent variables. When extrapolated, it forecasts values of the dependent variable based on the strength and direction of relationship with the independent variables. Regression has been used in the fields of econometrics and law (Sykes, n.d.), in improving students’ performance (Zakhem, Khair, & Moucary, 2011), and in setting health targets (Fukada, Nakamura, & Takano, 2002).

In performing regression analysis, data-points are first plotted in a scatter plot. Scatter plots allow the user to visually identify relationships between variables. The correlation coefficient is then found. The correlation coefficient is a standardized number between -1 and 1 that describes the strength and direction of relationship between variables. A correlation coefficient of 1 or close to 1 shows that there is a strong positive relationship between the variables. A correlation coefficient of -1 or close to -1 shows that there is a strong negative relationship between the variables. A correlation coefficient of 0 or close to 0 shows that there is no or weak relationship between the variables. Using the least-squares method, the regression line and the regression equation is found. The regression equation allows us to forecast or predict the dependent variable. Regression can be split into two types: linear and non-linear. Under each type, there are simple and multiple. In this paper, we will be using the multiple linear regression model.

Multiple linear regression allows for examining how multiple independent variables are related to a dependent variable (Higgins, 2005). R (also known as the multiple correlation coefficient) is used in multiple linear regression. It represents the strength and direction of relationship amongst a combination of variables. This differs from the simple linear correlation coefficient which only compares between two variables. In knowing the strength of relationship amongst all the variables, the multiple regression formula is formed. The multiple regression formula is as follows:

**Y = a + b1X1 + b2X2 + … bkXk**

Y – The value of the dependent variable

a – The intercept

b – The change in Y for each incremental change in X

X – The value of the independent variable

## **Evaluating Staff Performance using Clustering**

 Staff performance evaluation is an important base for monitoring store performance. Identifying groups of good performing and poor performing staff will help managers identify the percentage of staff that fall into good or poor performing groups. However, evaluation using just one indicator (eg: average sales) does not show other key characteristics of a good or poor performing staff. Clustering analysis helps solve this issue where multiple attributes can be taken into account, and used to cluster data-points that are similar. Islam & Haque (2012) describes clustering as an unsupervised data analysis method that classifies similar data into homogenous groups called clusters. Within each cluster its data-points are similar to one another, but the data-points are dissimilar between each cluster. Clustering can be categorized into two types - hierarchical and non-hierarchical. Examples of hierarchical clustering are single linkage, complete linkage, average linkage, median, and Ward. Examples of non-hierarchical clustering are k-means, k-medoids, and fuzzy clustering. In this paper, we will be using k-means clustering to analyze store staff performance.

K-means clustering analyses the Euclidean distance between data points and form K clusters of data points. It first assigns K number of data points to become centroids and calculate the Euclidean distance from the rest of the points (non-centroids) to these centroids. It then assigns non-centroids to be clustered with the nearest centroids. The process is reiterated for a fixed number of times for K clusters to try to maximize the inter-cluster distance (distance between each of the K centroids) and the intra-cluster distance (distance between non-centroids and their cluster centroids). In academic literature, k-means clustering has been used to in crime analysis (Agarwal, Nagpal, & Sehgal, 2013), predicting students’ performance (Oyelade, Oladipupo, & Obagbuwa, 2010), and improving students’ performance (Islam & Haque, 2012).

# **The Company - Teppei Syokudo**

Teppei Syokudo is a Japanese F&B chain, operating under the umbrella of the famous Teppei Japanese Restaurant. It operates as a takeaway kiosk concept, and specializes in Kaisendon, a bowl of barachirashidon (rice with raw fish). It offers other dons (rice bowls), tonkatsu cutlets (fried pork), tempura (fried dishes) and sashimi salads. The takeaway kiosk mainly targets lunchtime crowds, especially upmarket businessmen and women who are looking to grab a quick, healthy, and quality lunch. There are currently 4 outlet locations - Millenia Walk, Takashimaya, Republic Plaza and Ion Orchard. For this paper, we will only be analyzing data from the Millenia Walk outlet.

Each outlet has its own shop manager, cashier, packer, and kitchen staff. During peak hours, the outlets may have up to 5 staff to manage demand. Some staff work on a part-time basis. The store manager decides the staff roles for the day, hence some staff may be assigned different roles on different days. However, the job rotation usually occurs between part-timers, and rotation between cashier and packer role.



*Figure 1.* Sales Process

Looking at the sales process, it would seem that the cashier has the greatest opportunity to affect customers’ order decisions, leading to the day’s sales. It would seem that the kitchen staff and packer contribute to the speed of service provided to the customer. This may affect customers’ future decision in purchasing from Teppei Syokudo and such decisions would not be reflecting in the current day’s sales. The store manager oversees operations, but generally does not get involved in the sales process. From this observation, we hypothesize that the cashier is the only one who would be able to affect sales dollars. If this is the case, it would make the most sense to attribute sales only to staff who are involved in the cashier role.

## **Problem Definition**

Most F&B businesses measure their overall store performance through customer volume, sales, and profits. However, store managers are unsure as to which factors to focus on to increase overall sales. One can choose to focus on drinks sales or set meal sales, but which is more important? Currently, Teppei Syokudo measures store performance using the following weekly Key Point Indicators (KPIs):

Table 1

*Existing KPIs*

|  |  |
| --- | --- |
| KPIs | Definition |
| Drink Sales | Weekly sales of drinks |
| Drink Sales % | % of drinks sales / total sales |
| Set Sales | Weekly sales of sets |
| Set Sales % | % of set sales / total sales |
| Total Sales | Total sales |
| Working hour | Total staff working hours in a week |
| Productivity | Total sales / working hour |

Using multiple linear regression, we evaluate the existing KPIs to identify whether these KPIs are relevant to the overarching target – total sales. At the same time, we have identified several other KPIs that may be relevant to total sales and can be considered implementing.

Table 2

*New KPIs to evaluate*

|  |  |
| --- | --- |
| KPIs | Definition |
| Sales/Customer | Weekly sales $ per customer |
| Fried Item Sales | Weekly sales of fried items |
| Fried Item % | % of fried item sales / total sales |
| Onigiri Sales | Weekly sales of onigiri |
| Onigiri % | % of onigiri sales / total sales |

 A more controllable factor to driving store sales is staff performance. When staff upsells and cross-sells, there is increased probability for customers to spend more. Staff performance can hence be evaluated based on various factors that lead to overall sales, and managers will be able to identify groups of good performing or poor performing staff. In order to evaluate overall staff performance based on multiple factors, we use k-means clustering and cluster the staff based on the above mentioned existing and new store KPIs. It can then be said that staff who perform well on several indicators that are related to overall sales, do contribute positively to overall store sales.

# **Data Analysis Methodology**

 This section is split into two parts: (1) Evaluating existing KPIs and establishing new KPIs relating to store performance using multiple linear regression, and (2) Evaluating staff performance using existing and new store KPIs using k-means clustering. Within each part, we discuss (1) the data preparation involved in the analysis, (2) the analysis methodology, and (3) the analysis results and its implications.

## **Evaluating and Establishing KPIs using Multiple Linear Regression**

### **Data Preparation**



*Figure 2.* Data Preparation

732 hourly observations of Teppei Syokudo’s existing KPIs were prepared in Excel in the above format. We then carry on the rest of the data preparation and analysis in SAS® JMP.

Before multiple linear regression is conducted, outliers that might skew the relationship between X and Y variables must be removed using the Outlier Analysis function. We select and exclude the outliers that are identified, using the “Select rows” and “Exclude rows” function. In this case, outliers are defined as Q = 3 times the interquartile range past the 10% and 90% quartile.



*Figure 3.* Removing Outliers

### **Analysis Methodology.** Multiple linear regression can then be conducted in JMP using the Fit Model function by selecting the dependent Y variable and the independent X variables that will be tested. In this case, we will be testing the relevance of Teppei Syokudo’s current KPIs on its store sales.



*Figure 4.* Fit Model

**Analysis Results and Implications.** The regression results can be seen in the Regression report.



*Figure 5.* Actual by Predicted Plot

Table 3

*Fit Model Results*

|  |  |
| --- | --- |
| R Square | 0.97 |
| R Square Adj | 0.97 |
| Root Mean Square Error | 36.16 |

Table 4

*Parameter Estimates*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Term | Estimate | Std Error | t Ratio | Prob>|t| |
| Intercept | -356.148 | 11.684 | -30.480 | <.0001 |
| Labour Productivity | 2.848 | 0.047 | 60.880 | <.0001 |
| Total number of Staff Hours | 123.229 | 3.111 | 39.610 | <.0001 |
| Set Sales | 0.110 | 0.043 | 2.550 | 0.011 |
| Set Sales % | -32.046 | 12.589 | -2.550 | 0.011 |
| Drinks Sales | 1.777 | 0.581 | 3.060 | 0.002 |
| Drinks Sales % | -383.334 | 131.399 | -2.920 | 0.004 |

All of the terms used resulted in a Prob>|t| of <0.05. Therefore, the above regression shows that all of Teppei Syokudo’s current KPIs are relevant in affecting the store’s financial performance. The equation below illustrates the relationship of the factors to sales:

**Sales = 2.848(Labour Productivity) + 123.229(Total number of Staff Hours) + 0.110(Set Sales) – 32.046(Set Sales %) + 1.777(Drink Sales) – 383.334(Drink Sales %) – 356.148(intercept)**

Therefore, we conclude that Teppei’s current KPIs are relevant in affecting the store’s financial performance and have the following implications. Shop sales have a positive relationship with labour productivity, total number of staff hours, set sales and drink sales, but have a negative relationship with Set Sales % and Drink Sales %.

### **“Time of the day” effect.** One main phenomenon that F&B businesses usually face is the “Time of the Day” effect. Sales are at the highest during lunch and dinner, and there is an idle period between the two peak periods. We realized that the relationship between the KPIs and Store sales might vary at different times of the day due to the “Time of the Day” effect. We checked our data set and confirmed that the phenomenon does occur in our data set as well. To remove the Time of the Day effect on the regression model, we did three separate regression analyses for the three main periods of the day (Lunch Peak, Idle and Dinner Peak).

Table 5

*Time Periods*

|  |  |  |
| --- | --- | --- |
| Time of Day | Start | End |
| Lunch | 11am | 2pm |
| Idle | 2pm | 5pm |
| Dinner | 5pm | 8pm |

### **Additional KPIs.** We decided to add 2 different categories of KPI to reflect the effect of cross-selling as well as upselling. Thus for cross-selling, we decided that another dimension to determine cross-selling was to include the sales of fried items as well as onigiri, instead of just drinks sales. To reflect the dimension of upselling, we included the KPI of sales per customer (sales dollar / number of customers). This is because if the cashier undertakes upselling, each customer may end up spending more. Thus we seek to observe the sales per customer to determine if that does indeed occur.

#### **Lunch Peak Analysis**

**Existing KPIs**

Table 6

*Lunch Peak Results*

|  |  |
| --- | --- |
| RSquare | 0.977 |
| RSquare Adj | 0.976 |
| Root Mean Square Error | 38.517 |
| Mean of Response | 459.692 |

Table 7

*Lunch Peak Parameter Estimates*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Term** | **Estimate** | **Std Error** | **t Ratio** | **Prob>|t|** |
| **Intercept** | -433.4471 | 27.79 | -15.6 | **<.0001** |
| **Drinks Sales** | 1.7918 | 1.44 | 1.25 | 0.2143 |
| **Drinks Sales %** | -789.4109 | 520.19 | -1.52 | 0.1307 |
| **Set Sales** | 0.0005 | 0.08 | 0.01 | 0.9955 |
| **Set Sales %** | -3.2129 | 34.67 | -0.09 | 0.9263 |
| **Total number of Staff Hours** | 143.4834 | 6.08 | 23.59 | **<.0001** |
| **Labour Productivity** | 3.0762 | 0.09 | 32.77 | **<.0001** |

By looking at the above regression of Lunch Peak observations, it seems that out of the existing KPIs, only Labour Productivity and Total number of Staff Hours have a significant effect on Shop Sales Prob>|t| is <0.05.

Regression equation (Existing KPIs):

Sales = 3.076 (Labour Productivity) + 143.483(Total number of Staff Hours) - 433.447 (intercept)

**Existing and New KPIs**

Table 8

*Lunch Peak Parameter Estimates with New KPIs*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Term** | **Estimate** | **Std Error** | **t Ratio** | **Prob>|t|** |
| **Intercept** | -418.8419 | 36.91697 | -11.35 | **<.0001** |
| **Sales/Customer \*** | 1.2100197 | 0.641145 | 1.89 | 0.0606 |
| **Labour Productivity** | 2.9492286 | 0.113394 | 26.01 | **<.0001** |
| **Drinks Sales** | 1.7320596 | 1.473903 | 1.18 | 0.2414 |
| **Drinks Sales %** | -732.8349 | 527.6823 | -1.39 | 0.1665 |
| **Fried Items Sales \*** | 0.3034475 | 0.640934 | 0.47 | 0.6364 |
| **Fried Items Sales % \*** | -293.1096 | 265.8736 | -1.1 | 0.2717 |
| **Set Sales** | 0.0192657 | 0.083642 | 0.23 | 0.8181 |
| **Set Sales %** | -18.41525 | 34.39256 | -0.54 | 0.593 |
| **Onigiri Sales \*** | 0.9430865 | 0.546785 | 1.72 | 0.0862 |
| **Onigiri Sales % \*** | -423.6288 | 221.2719 | -1.91 | 0.057 |
| **Total number of Staff Hours** | 139.22444 | 6.942599 | 20.05 | **<.0001** |

**\* New KPIs**

After adding in the new KPIs and conducting the regression again, it seems that the new KPIs do not significantly affect Store Sales and the relationship remains.

Regression equation (With New KPIs):

**Sales = 2.949 (Labour Productivity) + 139.224(Total number of Staff Hours) - 418.842 (intercept)**

***Idle Period Analysis***

**Existing KPIs**

Table 9

*Idle Period Parameter Estimates*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Term** | **Estimate** | **Std Error** | **t Ratio** | **Prob>|t|** |
| Intercept | -149.1064 | 10.70563 | -13.93 | **<.0001** |
| Drinks Sales | 2.1158991 | 0.730485 | 2.9 | **0.0041** |
| Drinks Sales % | -217.5173 | 90.74541 | -2.4 | **0.0172** |
| Set Sales | 0.1845268 | 0.068123 | 2.71 | **0.0072** |
| Set Sales % | -20.47577 | 9.72292 | -2.11 | **0.0361** |
| Total number of Staff Hours | 56.04762 | 2.872187 | 19.51 | **<.0001** |
| Labour Productivity | 2.4736479 | 0.070854 | 34.91 | **<.0001** |

By looking at the above regression of idle period observations, it shows that all of Teppei Syokudo’s current KPIs are relevant in affecting the store’s financial performance during idle periods.

Regression equation (Existing KPIs):

**Sales = 2.474 (Labour Productivity) + 56.048(Total number of Staff Hours) + 0.185 (Set Sales) – 20.476(Set Sales %) + 2.116(Drink Sales) – 217.517(Drink Sales %) – 149.106 (intercept)**

**Existing and New KPIs**

Table 10

*Idle Period Parameter Estimates with New KPIs*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Term** | **Estimate** | **Std Error** | **t Ratio** | **Prob>|t|** |
| Intercept | -129.9926 | 11.51499 | -11.29 | **<.0001** |
| Sales/Customer \* | 1.0209688 | 0.234374 | 4.36 | **<.0001** |
| Labour Productivity | 2.0995678 | 0.081388 | 25.8 | **<.0001** |
| Drinks Sales | 2.1381534 | 0.662879 | 3.23 | **0.0014** |
| Drinks Sales % | -204.2243 | 82.63846 | -2.47 | **0.0141** |
| Fried Items Sales \* | 0.78384 | 0.186307 | 4.21 | **<.0001** |
| Fried Items Sales % \* | -93.52678 | 22.17662 | -4.22 | **<.0001** |
| Set Sales | 0.2526243 | 0.062651 | 4.03 | **<.0001** |
| Set Sales % | -30.5382 | 8.929709 | -3.42 | **0.0007** |
| Onigiri Sales \* | 0.9877765 | 0.267779 | 3.69 | **0.0003** |
| Onigiri Sales % \* | -79.98931 | 32.35192 | -2.47 | **0.0141** |
| Total number of Staff Hours | 49.35219 | 2.776793 | 17.77 | **<.0001** |

**\* New KPIs**

After adding in the new KPIs and conducting the regression again, it seems that the new KPIs are also relevant, alongside existing KPIs, in affecting the store’s financial performance during idle periods.

Regression equation (With New KPIs):

**Sales = 2.100 (Labour Productivity) + 49.352(Total number of Staff Hours) + 0.253 (Set Sales) – 30.538(Set Sales %) + 2.138(Drink Sales) – 204.224(Drink Sales %) +1.021 (Sales/Customer) + 0.784 (Fried Items Sales) - 93.527(Fried Items Sales %) + 0.988 (Onigiri Sales - 79.990 (Onigiri Sales %) – 129.992 (intercept)**

*Dinner Peak Analysis*

**Existing KPIs**

Table 11

*Dinner Peak Parameter Estimates*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Term | Estimate | Std Error | t Ratio | Prob>|t| |
| Intercept | -326.344 | 21.95146 | -14.87 | **<.0001** |
| Drinks Sales | 4.4393912 | 0.954387 | 4.65 | **<.0001** |
| Drinks Sales % | -1512.629 | 328.594 | -4.6 | **<.0001** |
| Set Sales | 0.2651555 | 0.059386 | 4.46 | **<.0001** |
| Set Sales % | -90.43372 | 22.78762 | -3.97 | **<.0001** |
| Total number of Staff Hours | 134.12183 | 5.423123 | 24.73 | **<.0001** |
| Labour Productivity | 2.3340819 | 0.067443 | 34.61 | **<.0001** |

By looking at the above regression of dinner peak period observations, it shows that all of Teppei’s current KPIs are relevant in affecting the store’s financial performance during idle periods.

Regression equation (Existing KPIs):

**Sales = 2.334 (Labour Productivity) + 134.122(Total number of Staff Hours) + 0.265(Set Sales) – 90.433(Set Sales %) + 4.439(Drink Sales) – 1512.629(Drink Sales %) – 326.344 (intercept)**

**Existing and New KPIs**

Table 12

*Dinner Peak Parameter Estimates with New KPIs*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Term | Estimate | Std Error | t Ratio | Prob>|t| |
| Intercept | -229.4184 | 22.58214 | -10.16 | **<.0001** |
| Sales/Customer | 0.7765645 | 0.291031 | 2.67 | **0.0082** |
| Labour Productivity | 1.9303612 | 0.074754 | 25.82 | **<.0001** |
| Drinks Sales | 3.7648659 | 0.829602 | 4.54 | **<.0001** |
| Drinks Sales % | -1293.542 | 285.5461 | -4.53 | **<.0001** |
| Fried Items Sales | 1.691182 | 0.280895 | 6.02 | **<.0001** |
| Fried Items Sales % | -533.1355 | 95.70708 | -5.57 | **<.0001** |
| Set Sales | 0.3605908 | 0.052742 | 6.84 | **<.0001** |
| Set Sales % | -127.5484 | 20.30767 | -6.28 | **<.0001** |
| Onigiri Sales | 3.0829225 | 0.598456 | 5.15 | **<.0001** |
| Onigiri Sales % | -945.8598 | 175.1345 | -5.4 | **<.0001** |
| Total number of Staff Hours | 111.2329 | 5.408912 | 20.56 | **<.0001** |

After adding in the new KPIs and conducting the regression again, it seems that the new KPIs are also relevant, alongside existing KPIs, in affecting the store’s financial performance during dinner peak periods.

Regression equation (With New KPIs):

Sales = 1.903 (Labour Productivity) + 111.233(Total number of Staff Hours) + 0.361 (Set Sales) – 127.548(Set Sales %) + 3.765(Drink Sales) – 1293.542(Drink Sales %) +0.777 (Sales/Customer) + 1.691 (Fried Items Sales) - 533.136(Fried Items Sales %) + 3.083 (Onigiri Sales - 945.860 (Onigiri Sales %) – 229.418 (intercept)

**Neglect “time of day” effect.** Although the current KPIs are relevant in explaining store sales as a whole, our findings show that the KPI has neglected to consider the time of day effect. As observed, the existing and new KPIs are relevant in explaining the store sales. However, when we break it down to account for the time of day effect, we realise that for lunch time analysis, only 2 KPIs, namely labour productivity and total number of staff hours, were relevant in explaining store sales. For idle and dinner times, all KPIs were relevant in explaining store sales. These leads us to 2 conclusions. Firstly, that management needs to be aware that their KPIs at lunch time differ from other times of the day. Secondly, that management can explore other factors to include in their KPI that could potentially explain the store sales at lunch time compared to other times of the day.

**Negative drink % relationship to store sales**. Throughout our analysis, we noticed a negative relationship between store sales and drink % sales. Management has highlighted their concerns that drink sales % are not increasing and that it could negatively impact the overall store sales and thus hurt profit. Thus in the past, when there was a dip in drink sales %, management would think that something is going wrong. However, our findings show that management should not be alarmed because of the presence of this negative relationship.

**Existing and New KPIs - keep or remove?**

In our analysis, we explored existing and new KPIs. Based on the findings, it shows that both new and existing KPIs generally result in the same conclusion. This means that in instances where existing KPIs are significantly relevant, new KPIs are as well. However, to consider the adding of new KPIs, we explore the adjusted Rsquare to ensure that we are not overfitting the data. The adjusted Rsquare of existing KPIs are 0.976 and the adjusted Rsquare of existing and new KPIs are 0.971. Since the Rsquare is reduced by 0.005 with additional KPIs added, this indicates that the new KPIs improves the model by less than expected by chance. Therefore, after exploring other items that can contribute to store sales, we conclude that it will be best for management to not include the new KPIs we have proposed. This results also shows that the current explanatory power of existing KPIs are sufficient and useful in explaning store sales.

**Existing and New KPIs - overall** (do not account for “time of day” effect)

|  |  |
| --- | --- |
| RSquare | 0.970946 |
| RSquare Adj | 0.970501 |
| Root Mean Square Error | 36.58327 |
| Mean of Response | 301.8778 |
| Observations (or Sum Wgts) | 730 |

# **Staff Performance using Clustering**

## **Data preparation**



We collected data that specified the cashier for each hour for all 730 hourly observations.



We then joined the Cashier variable with the rest of the hourly KPIs.

## Analysis Methodology

To ascertain whether different cashiers have an effect on the KPIs, as well as the store’s financial performance, we conducted regressions by time of the day between the independent variable (the cashier) and the dependent variables (the rest of the KPIs)



|  |  |  |  |
| --- | --- | --- | --- |
| **Time of Day** | **Effect** | **Response** | **Prob > F** |
| Dinner | Cashier | Sales/Customer | **0.0073** |
| Dinner | Cashier | Labour Productivity | **0.0495** |
| Dinner | Cashier | Set Sales | **0.0002** |
| Dinner | Cashier | Set Sales % | **0.0001** |
| Dinner | Cashier | Total number of Staff Hours | **0.0234** |
| Idle | Cashier | Labour Productivity | **0.0001** |
| Idle | Cashier | Set Sales | **0.0019** |
| Idle | Cashier | Total number of Staff Hours | **0.0366** |
| Idle | Cashier | Sales | **0.0001** |
| Lunch | Cashier | Sales/Customer | **0.0001** |
| Lunch | Cashier | Labour Productivity | **0.0001** |
| Lunch | Cashier | Fried Items Sales % | **0.0169** |
| Lunch | Cashier | Set Sales | **0.0007** |
| Lunch | Cashier | Onigiri Sales | **0.0043** |
| Lunch | Cashier | Total number of Staff Hours | **0.0001** |
| Lunch | Cashier | Sales | **0.0001** |

Significant relationship between Cashier and KPIs by time of day

The significant responses were clustered for each time of the day using K-means.

The best K values were chosen based on their CCC performance measures.

**Lunch**

|  |  |  |  |
| --- | --- | --- | --- |
| Method | NCluster | CCC | Best |
| K-Means Clustering | 3 | -4.2821 |  |
| K-Means Clustering | 4 | -1.9218 | Optimal CCC  |
| K-Means Clustering | 5 | -3.5325 |  |

**Idle**

|  |  |  |  |
| --- | --- | --- | --- |
| Method | NCluster | CCC | Best |
| K-Means Clustering | 3 | -5.5284 |  |
| K-Means Clustering | 4 | -3.6395 |  |
| K-Means Clustering | 5 | -3.4114 |  |
| K-Means Clustering | 6 | -1.9825 |  |
| K-Means Clustering | 7 | -1.8663 | Optimal CCC  |

**Dinner**

|  |  |  |  |
| --- | --- | --- | --- |
| Method | NCluster | CCC | Best |
| K-Means Clustering | 3 | -3.0935 |  |
| K-Means Clustering | 4 | -2.4146 | Optimal CCC  |
| K-Means Clustering | 5 | -3.5009 |  |

Clustering was performed for the optimal number of k for each time of the day. As seen from the table, each cluster performs uniquely in each KPI. Each cashier‘s working hours are segmented into different clusters to determine whether there are cashiers who can be consistently found in a certain cluster.

**Clusters of KPIs for Lunch Peak**

****

**Implications.** Cashiers’ performance under each KPI varies greatly under each KPI and it is unlikely that cashiers would affect the KPIs. Hence, further studies need to be done to ascertain other independent variables that might drive the relevant KPIs.

# **References**

Agarwal, J., Nagpal, R., & Sehgal, R. (2013). Crime analysis using k-means clustering.

*International Journal of Computer Applications*, *83*(4), 1–4. doi:10.5120/14433-2579

Fukuda, Yoshiharu, Nakamura, Keiko, & Takano, Takehito. (2002). A combination of an

extrapolation method and a benchmark method to develop quantitative health targets for Japan. *Health Policy,* *61*(2), 201-212.

Higgins, Jim. (2005). *The Radical Statistician.* Retrieved from

<http://www.biddle.com/documents/bcg_comp_chapter4.pdf>

Islam, H., & Haque, M. (2012). An approach of improving Student’s academic performance by

using k-means clustering algorithm and decision tree. *International Journal of Advanced Computer Science and Applications*, *3*(8), 146–149. doi:10.14569/ijacsa.2012.030824

Oyelade, O. J, Oladipupo, O. O, & Obagbuwa, I. C. (2010). Application of k-Means Clustering

algorithm for prediction of Students’ Academic Performance. *International Journal of Computer Science and Information Security,* *7*(1), 292-295.

Sykes, Alan O. (n.d.). An Introduction to Regression Analysis.

<http://www.law.uchicago.edu/files/files/20.Sykes_.Regression.pdf>

Walid Zakhem, Marie Khair, & Chady El Moucary. (2011). Improving Students Performance

Using Data Clustering and Neural Networks in Foreign-Language Based Higher Education. *International Journal of ACM Jordan,* *2*(1), International Journal of ACM Jordan, 01 September 2011, Vol.2(1).