

APPROACHES TO CUSTOMER DATA MANAGEMENT AND KNOWLEDGE DISCOVERY IN SMALL RETAIL ANALYTICS: THE CASE OF A TINY COFFEE SHOP IN BANGKOK, THAILAND

Pakawat Paiboon*

Business Analytics and Intelligence Program, School of Applied Statistics
National Institute of Development Administration (NIDA)
Bangkok, Thailand
Email: pakawat.pai@stu.nida.ac.th

Akkaranan Pongsathornwiwat

Logistics Management Program, School of Applied Statistics
National Institute of Development Administration (NIDA)
Bangkok, Thailand
Email: akkaranan@as.nida.ac.th

ABSTRACT

Although coffee shop is the highest competition comparing to others in Thailand for the past four years, only 10% of all new restaurants are success. To increase a success rate, data as an important resource needs to be significantly used to increase sales and maintain advantages. However, complications of data processing and lacking analytical models are common challenges to small businesses management as it requires huge investment on data management system. This study proposed data management framework as economically solutions to processing existing customer data into actions. Data from the tiny coffee shop in Bangkok as our case study are used for illustration. In data management, existing data were designed to combine all daily transactions into one structural file by using python script and use it as a data source for reporting. As of reporting, LINE chatbot was selected as a channel to send quick reports and analytic results of co-occurrence items, and Dashboards with interactive filtering ability via Power BI are employed for detailing sale information. These approaches can help owners to have deep understanding on sale and customer data which leads to better profits, and be applied to any retails that face with similar issues of customer analytics.

KEYWORDS: 1) CUSTOMER DATA MANAGEMENT 2) APPLIED DATA MINING
3) CHAT BOT 4) RETAIL ANALYTICS 5) CASE STUDY

1. Introduction

70,149 restaurants were newly opened in Thailand during 2019. However only 10% of those restaurants are successful during the past three years. Interestingly, Bangkok has the highest number of newly restaurants opened in Thailand. A secondary survey from Wongnai's market research in 2020 shows that there are 18,000 new restaurants out of 70,149 (Wongnai, 2020). In term of restaurant types, a coffee shop is the most popular ones among all types in Thailand, supported by 17,000 of new coffee opened in 2019 both franchises and new own brand which is represented 81% of growth rate, comparing to 2018 (Wongnai, 2020).

According to 90% of failures of new restaurants and the highest competitive environment of coffee shop, business owners target at increasing sale amounts while reducing all possible costs with available resources without additional costs in order to prolong survival rates (Suad S. Mohamud, et al., 2017). As suggested in the literature, the highest value of their available resources is data, for instance, sale data and transactions. By utilizing existing daily sale data, business owners can better insightfully understand on the trends of popular menu which one is the rising star or the beast (Pengfei W., et al., 2014). This leads the companies to gain more promotion opportunities and competitive advantages. One of the most important data is the sale data from Point-of-Sales (POS) database that can enhance firms to deeply find out their customer behavior. With mentioned benefits, coffee-shop owners can use the right information for better profit optimizations. The advantages of customer analytics have been proved for the big firms where have enough funding (Ranjit, 2009). However, there is little knowledge of how small and medium-sized enterprises (SMEs), especially in Thailand, apply the data analytic and management effectively. This is the key to our study that will go through.

The objective is to propose the approaches to customer data management and knowledge discovery in small retail shops. The study case for this study is a small café and bistro in Bangkok name "WAAK". WAAK is a small café and bistro shop is facing similar problems with other small business due to small investment such as lack of financial liquidity, not good enough in store management, and especially lack of effectively use their store and customer data. The data for development is POS data such as transaction and inventory. The problem is that a basic capability of POS service usually provides only daily sale information sending through email in "csv" format. Based on those available data format led to inefficiently use and transform any data's insight to action due to complication of data structure, integration, and processing also lack of data analytics model together within platforms that always directly connect business owners to their data. Those all issues are root cause of lower opportunity for increasing profits through up-selling, cross-selling as well as cost reduction through inventory and resources optimization.

Thanks to a scant of limited capability on data management and data analytics based on small system's investment. All those challenges stand for the opportunities that aim to suggest small retail businesses to effectively use their data for better planning and execution. By creating data management and automated reporting system in order to help transforming data into action without additional cost. This is the main contribution of our study.

2. Literature Review

2.1 Data Management (Transaction) and Data Structure

Proper data management leads to clear insight and proper actions for profitability optimization of SME. There are a ton of services and software for data management especially for transaction sale data. However, those service also leads to significant additional cost with some unnecessary functions to business (Deli Zhang, 2016) that

because of business's uniqueness. Even though same business type, each of coffee shop is different from each other based on their theme, service quality, and location those leading to their customer, best seller menu, etc. Any of those impacts to data insight which leading to action to improve business profitability. Shouhong W. and Hai W. (2020) also stated in the similar way that strategic use of data, data requirements, and knowledge products are the major constructs of big data for small business those also lead to design of structure data that based on need of use.

2.2 Efficiency Data Reporting (Chatbot & Dashboard)

Now a day in digital world, it provides more comfortable also effective using any data. As same as in business, when compared with a manual system, the use of digital applications is more effective and efficient in business management (Rahmatullah, Inanna, Sahade, et al., 2020). Efficiently access to insight of data with time limitation and proper action are the keys success of business that have many studies support such as following: Mohsen A. & Jeremy W. (2019) have stated that the digital revolution is helping organizations transform their businesses to better engage and stay connected with their customers, suppliers, and employees. Fatma Chiheb, et al. (2019) have been studied on decision making model also mention that the analysis of data created day-to-day provides new information that has the potential to enhance the decision-making process, thus improve the quality of decisions and achieve a competitive advantage for organizations.

2.3 Association Rule & Promotion

Association rule learning is a rule-based machine learning method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using some measures of interestingness. Based on the concept of strong rules, Rakesh Agrawal, Tomasz Imieliński and Arun Swami introduced association rules for discovering regularities between products in large-scale transaction data recorded by point-of-sale (POS) systems in supermarkets (Wikipedia, 2020). This information can be used as the basis for decisions about marketing activities such as promotional pricing together with cost optimization also inventory optimization. Paul D.B. and Nada N.B., (2001) also support on benefit of application which stated that the major benefit of relationship marketing is the ability to make decisions based on their impact on customer equity. On Shaohui and Robert (2017)'s study also provide clear benefit of optimization models that their output of the model provides optimized prices, display and feature advertising planning together with sales and profit forecasts that generates accurate sales forecasts and increases category profits by approximately 17% and that including cross-item and cross-period effects is also valuable.

3. Research Methodology

Methodology for this study follows Cross-industry standard process for data mining, known as CRISP-DM that was developed in 1996 by incorporation of 3 companies which are Daimler Chrysler, SPS, and NCR (Colin Shearer, 2000). CRISP-DM is an open standard process model that describes common approaches used by data mining experts. It is the most widely used analytics model. The CRISP-DM breaks the process of data mining into six major phases. However, it was adjusted a bit by combining phase 5 and 6 together as showing in Table 1 to be more suitable applying for this study in small business case.

3.1 Business Understanding

The first step which is the most important step is business understanding, which is key of all proper following actions leading to good result. This step is combination of learning from WAAK’s business owner, being a customer, and digging into all available data especially transaction data in order to understand business’s strength, weakness and any business problem together with business owner. Keys strength of WAAK is relationship with customers lead to major of the customer is existing customer. For opportunity, main targets are maintaining existing customer also increasing new customer rate by understand customer behavior using sale data.

Table 1: Adjusted Data Analytics Lifecycle with concept and result in each phase

Phase	Concept & Methodology	Result
1. Business Understanding	learning from WAAK’s business owner, being a customer, and digging into all available data to understand business’s strength, weakness, and any business problem with business owner	Keys strength of WAAK is relationship with customers. For opportunities are maintaining existing customer and increasing new customer rate.
2. Data Understanding	Digging down into daily available 13 csv data. A file, sales by user report or daily transaction.csv, was selected to be a source of data understanding.	A new extracted data type is created from data understanding phase for data analytic (association-rule) requirement which is “Item Type”. Krapao (Stir fried basil) menu was separated from other food as it is popular menu (high sale amount) and put as another item type.
3. Data Preparation	Using data and understanding of business to develop one data structure that contain all information to be ready for dashboard and reporting tool. For data storage, select data storage that easy to access for reporting and analytic tools.	Daily transactions in csv format were transformed and combined to only one data structure including new data type, Item Type, by new development python script. For data storage, “Firebase” was selected to be data storage location.
4. Data Modeling and Visualization	Using new data structure to develop visualization tools that friendly use, easy to understand and transform any insights to action.	LINE chatbot on mobile is selected for quick reporting that will provide 6 different reports based on business owner request. For more detail of daily sale, Dashboard is selected another tool using Power BI.
5. Evaluation and Deployment	Tools deployment at the same time with feedbacks gathering from business owners is key for this phase.	Chatbot and Dashboard were adjusted along the way to match with business owner needs with no additional cost.

3.2 Data Understanding

Second step, data understanding, 13 csv sale information files from POS system have been attached in daily email which is source of available data. All those available files are in format that cannot immediately use even there is csv format due to a lot of space between cell and column data.

In term of valuable data, information from all 13 files is quite duplicate as all related to sale information that transformed to different display in 13 templates. So, by digging down into available data, a file, sales by user report or daily transaction.csv was selected to be a source of data understanding also used to design new data structure for WAAK.

An interesting data was observed during digging down to transaction data, Krapao menu (Stir fried basil) is popup from other food menu (high sale amount), figure 1 show bar chart of top sale item, which related to discussion with business owner that Krapao is the best seller of food menu. This led to separation of Krapao menu from other food in Item type (new extracted data type). Even Beer is higher than Krapao but it was excluded from the interesting item and groped in beer group.

3.3 Data Preparation

Third step is Data Preparation, daily transaction in csv format were transformed and combined to only one data structure that contain all information, ready for dashboard and reporting by new development python script for WAAK’s data loading and data

transforming. One additional data also created from data understanding phase for data analytic (association-rule) requirement which is “Item Type”. The data storage is also important that need to be location that easy to access for reporting and analytic tools. For this study “Firebase”, which is a mobile and web application development platform, was selected to be data storage location.

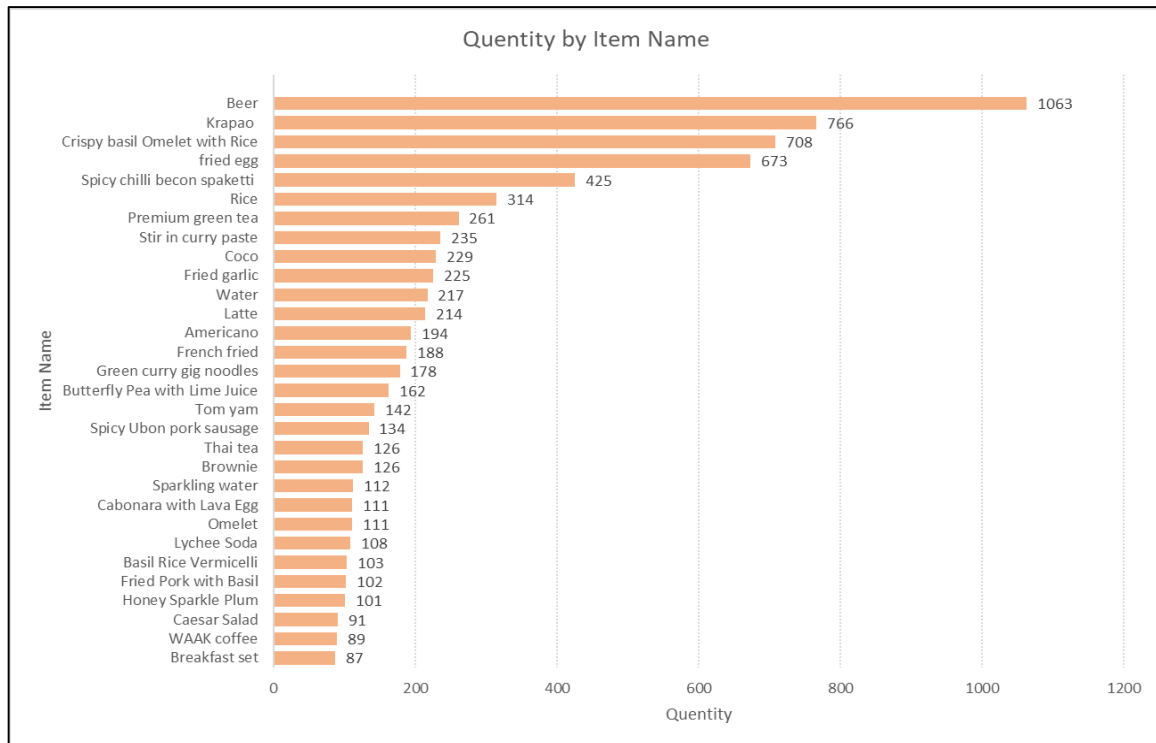


Figure 1: Bar chart show ranking of sale amount of each item.

3.4 Data Modelling and Visualization

Fourth step is Data analytics, Data visualization and Reporting, keys for this step are friendly use, easy to understand and transform any insights to action. For quick reporting whenever user want, chatbot on mobile is selected channel for connecting user to their data. LINE is not only the most famous instant communications application on mobile in Thailand but also a lot of available free platform and service to support conversational interfaces creation for chatbot especially Dialogflow those are the reasons to select LINE to be application for chatbot development in this study. Based on data understanding phase, information that related to daily sale amount and top item sale are the most popular that business owner needs to know by quick report that were designed and separated to 6 reports in chatbot.

However, they are some limitations of quick data reporting on mobile such as reporting space (small size and limited detail of information) and interactive filtering ability, so dashboard is selected to fulfill those limitations. Power BI was selected as a service tool for dashboard development in this study as it friendly use and many functions that match with WAAK information. The dashboards were designed to serves business owner need about detail daily sale information which are sale amount, sale quantity, sale by item also ability to filter and adjust the dashboards.

3.5 Evaluation & Deployment

As small business, study designed to combine phase 5 and phase 6 of CRISP-DM which are deployment the tools together with gathering feedbacks from business owners for tools adjustment along the way as no additional cost compare to big business scale. After been through all methodology steps using adjusted CRISP-DM, summary of data management and data reporting workflow for WAAK were summarized as figure 2 below. Daily data from email will be manually downloaded and processed together with existing data using python script. Json file format is one of product from processing process that need to be uploaded to Firebase. Once file is in Firebase, all data is ready for LINE chatbot and Dashboards to serve information to business owners.

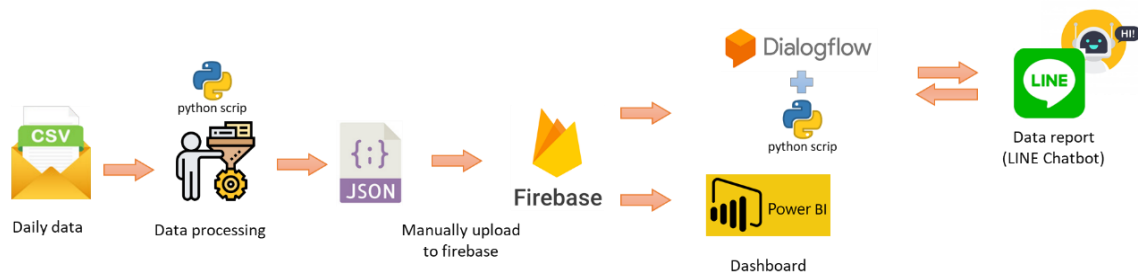


Figure 2: Data Management Solution Summary

4. Results

4.1 Data Management Solution

4.1.1 New Extracted Data

Item Type is a new data that was extracted from daily sale information based on phase of data understanding together with business owner’s knowledge and expert comments. It was separated to 12 types that consist of: Kao, Kaopao, Soup, Noodle, Salad, Appetizer, Coffee, Cake, Soda, Tea, Milk, and Beer. For Krapao, it was separated from other food as it is the bestseller menu also difference in term of sale amount from other menus.

Not only used for describing sale item but Item Type also used for association rule analysis. The analysis’s result will return pair of item type that popular occur in the same transaction, lead to cross-sale and up-sale opportunity by doing marketing campaign such as sale promotion.

4.1.2 New Data Structure

Daily 13 sale information in csv file format that normally stored in personal email system, rarely used by business owner, were transformed by python script to one structure in table combining all daily sale data. New data structure is containing bill number in each day, timestamp of bill, date of bill, type of sale (dine or take away), sale item name, sale item type, item price, quantity, and sale amount. Example of new data is as figure 3.

No.	Bill	Timestamp	Date	Type	Item Name	Item type	Item Price	Quantity	Sales
1	1012095	01/12/2019 13:06	01/12/2019	Dine In	ข้าวไข่กระทะเพรากลอบ	krapao	99	1	99
1	1012095	01/12/2019 13:06	01/12/2019	Dine In	มอคค่าช็อคโกแลตชิฟฟิน	coffee	99	1	99
2	1012096	01/12/2019 13:29	01/12/2019	Dine In	สิ้นใจโซดา	soda_beverage	79	1	79
2	1012096	01/12/2019 13:29	01/12/2019	Dine In	ข้าวไข่กระทะเพรากลอบ	krapao	99	1	99
3	1012097	01/12/2019 15:04	01/12/2019	Dine In	เค้กเวทเค้ก	cake	79	1	79
3	1012097	01/12/2019 15:04	01/12/2019	Dine In	ชาเอิร์ลเกรย์ Earlgey tea	tea	60	1	60
4	1012098	01/12/2019 15:57	01/12/2019	Dine In	{วรรค}โซดา	soda_beverage	65	1	65
4	1012098	01/12/2019 15:57	01/12/2019	Dine In	เส้นหมี่กะเพรา	krapao	65	1	65
4	1012098	01/12/2019 15:57	01/12/2019	Dine In	ไข่ดาว	etc	10	1	10
4	1012098	01/12/2019 15:57	01/12/2019	Dine In	ส่วนลด	etc	-6.5	1	-6.5
5	1012099	01/12/2019 17:30	01/12/2019	Dine In	นมหน้าสิ่ง	milk	85	1	85
6	1012100	01/12/2019 17:30	01/12/2019	Dine In	เอสเปรสโซ	coffee	50	1	50
6	1012100	01/12/2019 17:30	01/12/2019	Dine In	{วรรค}โซดา	soda_beverage	65	1	65

Figure 3: Example of new data structure

4.1.3 Data Storage

Firestore cloud data storage on web base was selected to be WAAK's data storage location for new data structure instead of email system. This help in term of data access for data reporting both chatbot and dashboard.

4.2 Data Reporting and Dashboard Visualization

4.2.1 LINE Chatbot

By observation and discussion with WAAK business owner to requirement of quick data reporting, 6 report types were created on chatbot to serves information by requesting with easy conversations on chat board of LINE as figure 4(a), the 6 reports type are consisting of:

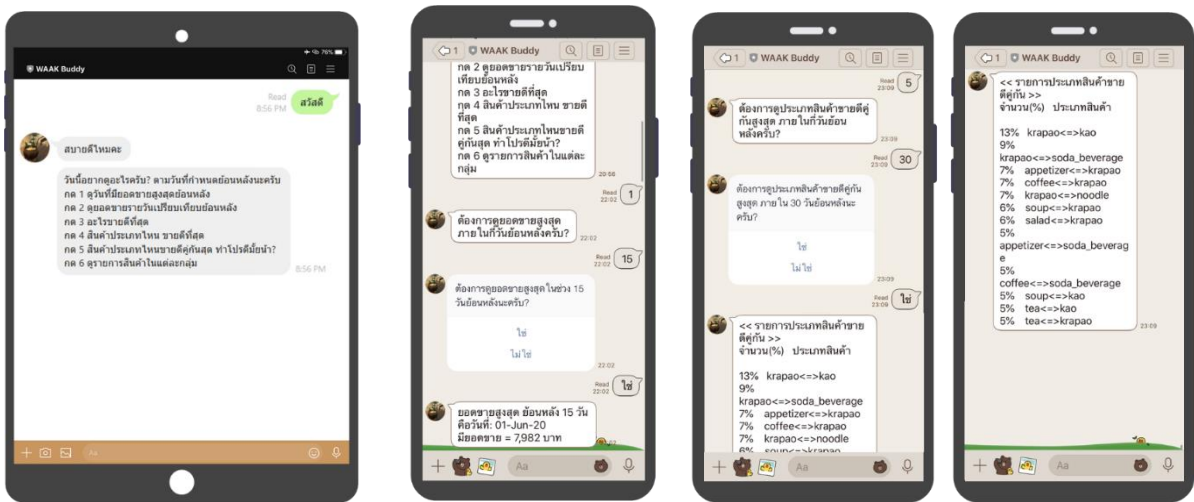


Figure 4: Examples of conversation on requesting information in LINE chatbot and maximum sale report (a): 6 report on Line chatbot, (b): example of Maximum sale amount and date during pass specific day report, (c, d): example of Top pair item type sale (Co-occurrence) report

4.2.1.1 Maximum sale amount and date during pass specific day (figure 4(b))

4.2.1.2 Daily sale amount during pass specific day

4.2.1.3 Top items sale

4.2.1.4 Top item type sale

4.2.1.5 Top pair item type sale (Co-occurrence)

The report required user to input number of day period to calculate and return support (%) and pair item type that ranking from higher support value based on minimum support that was set at 5% as figure 4(c,d). Only minimal information was included in this report by excluding other information such as confidence, consequence support, lift, leverage to be

simple for business owner understanding and applying. This is main benefit for sale optimization such as sale promotion.

4.2.1.6 Item list of item type

This report was created to help business owner easily access to list of items in each item type in order to identify item for future marketing action such as sale promotion based on co-occurrence report.

4.2.2 Dashboard via Power BI

Interactive filtering ability is required for detail sale information report. So, dashboard was selected to serve this ability via using Power BI. The sale information dashboards automatically pull information from Firebase using python script (same data source as the chatbot). 5 dashboards were created to deliver sale information in 3 different perspectives as following.

4.2.2.1 Overall sale information

Overall sale information dashboard will serve overall sale information that contain total sale, total customer number, quantity, and sale amount by item and by item type of specific interesting period that allow user to adjust, the example of this dashboard is showed in figure 5.

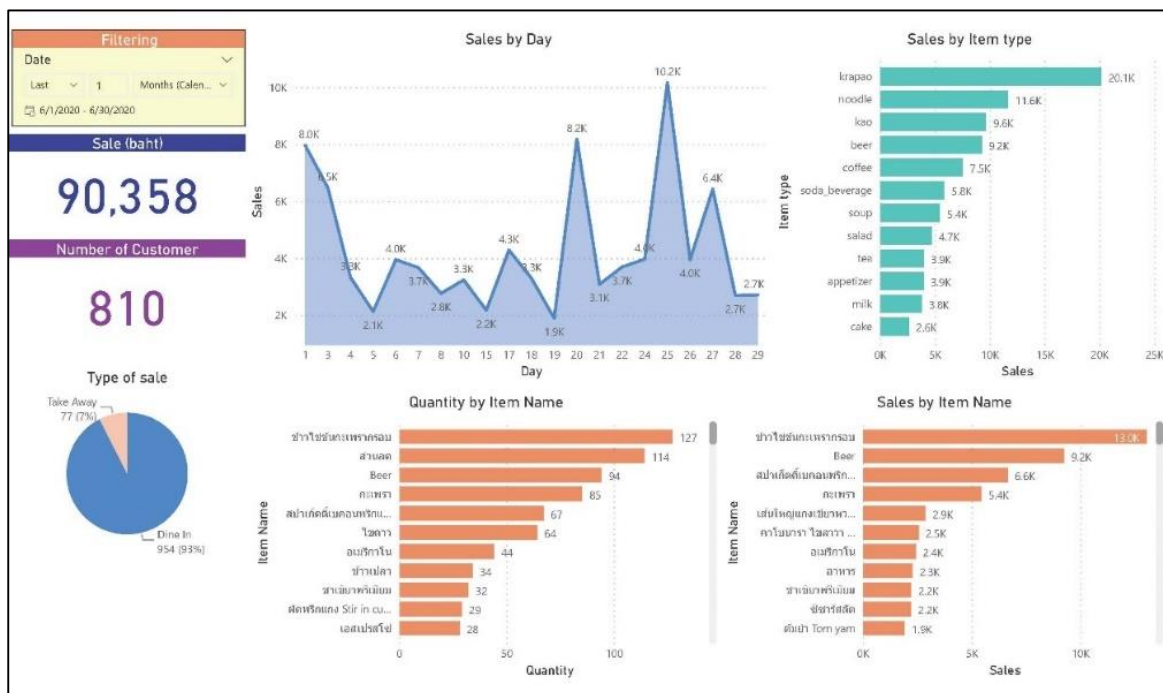


Figure 5: Example of overall sale information dashboard

4.2.2.2 Total sale amount by day and by Item type

2 total sale information dashboards provide total sale information in daily basis and total sale by Item type basis of specific interesting period that easily select by user.

4.2.2.3 Total sale divided into Item type in daily and monthly

2 total sale separated by Item type dashboards provide total sale information that separated into Item type in daily and monthly basis.

5. Discussions & Conclusions

5.1 Data Management

Combining daily sale information into one file is a big improvement for data management based on business owners' feedbacks that help them easily access to their sale data instead of using email system looking day by day in different email files as previous. Currently, business owners also have ability to use familiar program such as Microsoft Excel with new data structure to easily filter and dig down into their interesting data. Using cloud service to be data storage location is another benefit in terms of data accessibility also support any other application development in the future as well. In term of data management are met the study objectives that help business owner effectively use their data without any additional cost.

For improvement opportunities, daily new sale data still required manual download, run python script, and upload to Firebase cloud service. One opportunity to improve this process is automate system that can auto downloading attached csv files from email, processing, and auto uploading to cloud in order to get rid of these manual processes that will help on time management of business owner. Another point is additional information that could help using data more effective which are cost and benefit data of each item that can be generated from sale price information together with fix cost of store and variable cost of each item. That information will help to optimize more on marketing campaign. For cloud storage, back up location is also important to stability of application when the main server is down. So, by adding another file storage such as Heroku, Google, or Amazon as back up locations will support on application stability.

5.2 Data reporting and Customer analysis

By deployment both reporting tools quick report via LINE chatbot and Dashboards via Power BI are increasing efficiency of using data also transforming data's insight to action. Quick report via LINE chatbot is providing reports with immediately respond whenever business owner request also provides co-occurrence analysis that can be use as information for marketing campaign such as sale promotion. While Dashboards via Power BI are providing detail sale information reports that help business owner to evaluate and check store performance. With interactive filtering abilities, helps business owners easily access detail of sale data with many perspectives as they needed. In term of effectively used data are also met study objectives that help business owners understand more on their customer data also transform that to action without any additional cost.

In term of improvement opportunities, only text format is available in reports on LINE chatbot might not be interesting and comfortable enough, by adding dashboard in picture format might help on this potential issue together with adding more function into the chatbot based on business owner needed, via using menu template to select report instead of typing number. For customer analytics, membership system is very interesting that can help to get more insight of customer behaviors. Those data also can be used to get more accurate customer analytics such as collaborative filtering. By getting membership system can be both, develop system by our self or cooperate with existing membership service as a partner such as Privage, Gram Digital Publishing, or membership system from POS service. Cost and benefit will be main criteria for the service selection.

6. References

- Berger, Paul D. & Bechwati, Nada Nasr.2001. The allocation of promotion budget to maximize customer equity. **Omega**, Elsevier, vol. 29(1), pages 49-61.
- Bose, R. 2009. Advanced analytics: opportunities and challenges. **Industrial Management & Data Systems**, Vol. 109 No. 2, pp. 155-172.

- Colin Shearer.2000. The CRISP-DM model: The new blueprint for data mining. **Journal of Data Warehousing**, 5(4), page 13–22.
- Fatma Chiheb, Fatima Boumahdi, Hafida Bouarfa.2019. A New Model for Integrating Big Data into Phases of Decision-Making Process, **Procedia Computer Science**, vol. 151, 2019, Pages 636-642, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2019.04.085>.
- Mohsen Attaran & Jeremy Woods.2019. Cloud computing technology: improving small business performance using the Internet, **Journal of Small Business & Entrepreneurship**, 31:6, 495-519, DOI: 10.1080/08276331.2018.1466850
- Pengfei Wang, Jiafeng Guo, and Yanyan Lan. 2014. Modeling Retail Transaction Data for Personalized Shopping Recommendation. **In Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management (CIKM '14)**. Association for Computing Machinery, New York, NY, USA, 1979–1982. DOI:<https://doi.org/10.1145/2661829.2662020>
- Rahmatullah, Inanna, Sahade, Nurdiana, Fajriani Azis, and Bahri. 2020. Utilization of Digital Technology for Management Effectiveness Micro Small and Medium Enterprises. **International Journal of Scientific & Technology Research**, vol. 9(04), page1357-1362, ISSN 2277-8616
- Shaohui Ma and Robert Fildes. 2017. A retail store SKU promotions optimization model for category multi-period profit maximization, **European Journal of Operational Research**, Volume 260, Issue 2, Pages 680-692, ISSN 0377-2217
- Shouhong W. and Hai W.2020. (Big Data for Small Business: A Knowledge Management Model). **Northeast Decision Sciences Institute 2020 Annual Conference**: 244
- Suad Sheikh Mohamud, Gamal S. A. Khalifa, Abuelhassan E. Abuelhassan, and Sivakumar Kaliyamoorthy. 2017. Investigating the Antecedents of Coffee Shop Customers Behavioral Intentions In Kuala Lumpur. **International Journal on Recent Trends in Business and Tourism (IJRTBT)**, 1(4), 1-14.
- Ying Liu, Anthony Soroka, Liangxiu Han, Jin Jian, Min Tang.2019. Cloud-based big data analytics for customer insight-driven design innovation in SMEs, **International Journal of Information Management**, 2019, 102034, ISSN 0268-4012, <https://doi.org/10.1016/j.ijinfomgt.2019.11.002>.
- Zhang, Deli.2016. High-Performance Composable Transactional Data Structures.Electronic Theses and Dissertations. (2004-2019): 5069.
- Wongnai for business.2020. ส่งสถิติร้านอาหารปี 2019 ที่คนครัวต้องรู้ เปิดฟรี 70,000 ร้านแต่ปีตระหนาวรลือรอดเพียง 10%. Retrieved 15 June 2020 from <http://www.brandage.com/article/16993/Food-2019>
- Prasit Chulanutrakul.2019. **Tutorial: สร้าง Line Chat Bot ด้วย Dialogflow + Python (Heroku) + Firebase**. Retrieved 15 January 2020 from <https://medium.com/data-cafe-thailand/tutorial-%E0%B8%AA%E0%B8%A3%E0%B9%89%E0%B8%B2%E0%B8%87-line-chat-bot-E0%B8%94%E0%B9%89%E0%B8%A7%E0%B8%A2-dailogflow-python-heroku-firebase-part-i-36d51c6834a0>
- Witchapong Daroontham.2018. สร้าง **Line Chatbot ด้วย Dialogflow, Python, และ Firebase** ง่ายมาก!!!. Retrieved 18 January 2020 from <https://medium.com/datawiz-th/%E0%B8%AA%E0%B8%A3%E0%B9%89%E0%B8%B2%E0%B8%87-line-chatbot-%E0%B8%94%E0%B9%89%E0%B8%A7%E0%B8%A2-dialogflow-python-%E0%B9%81%E0%B8%A5%E0%B8%B0-firebase-%E0%B8%87%E0%B9%88%E0%B8%B2%E0%B8%A2%E0%B8%A1%E0%B8%B2%E0%B8%81%E0%B9%86-c4631c041848>