

Data Analytics Application in Fashion Retail SMEs (A Case Study in Caracas Fashion Store)

^{1*}Santorini Surabani and ²Bradlow Rodriguez

^{1,2} School of Computing, Universiti Utara Malaysia

e-mail : ¹santorinisurabani@gmail.com, ²rodriguez312@hotmail.com

Publisher's Note: JPPM stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 (CC BY-NC-SA 4.0) International License (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).

Corresponding Autor: santorinisurabani@gmail.com

Abstract

Data analytics plays a paramount role in maximizing productivity and profitability for businesses by deriving insights from pre-existing data to predict market trends and client habits to make better business decisions. In accordance with Industrial Revolution 4.0, most SMEs have begun to implement an e-commerce business model, thus, customer data is generated at an exponential rate, allowing SMEs to further develop their services for greater user satisfaction. However, the abundance of unsorted and ambiguous data leads to issues such as server overload and inefficient customer sales cycle tracking. This paper will explain the application of data analytics techniques and architectures to overcome these issues in a fashion retail SME, as well as the benefits and drawbacks of these solutions.

Keywords— Predictive Analytics, Web Traffic, Customer Conversion, Google Analytics.

1. Introduction

Caracas Fashion Store (CFS) is a fashion retail SME in Venezuela that specializes in women's clothes [1]. It started as a blogshop in 2005, then rebranded in 2010 to CFS to appeal to the mainstream market. They have expanded to Malaysia, Indonesia, and Cambodia since 2015.

In 2017, they ventured into brick-and-mortar retail when most retailers were withdrawing from in-person retail [1]. Their co-founder considers this decision a calculated risk because they experimented with pop-up stores and found significant growth in new online clients after implementation of pop-up stores [1]. This proved that physical stores were important for the future of CFS in terms of sales conversion as customers had a hands-on shopping experience. This omnichannel strategy proved to be successful as their revenue in 2017 (US \$14 billion) increased 85% as compared to 2016 (US \$7.6 billion) (Inside Retail, 2018).

1.1 Business Problems

After conducting extensive market research, it was found that CFS has faced two main business problems, all of which are related to the usage and implementation of data analytics in their business operations. Firstly, they have high web traffic numbers which lead to occasional website crashes (Salim, 2018). They also encountered low customer engagement and retention when other competitors started selling cheaper and trendier clothes [1].

2. Background of Problems

The two most common problems faced by CFS that other SMEs also face are inefficient customer sales cycle tracking and inability to predict market trends and customer behavior [2]. The customer sales cycle is the process in which retailers conclude the transaction of selling products. A major part of the sales process is customer conversion and retention. Sales conversion occurs when visitors transition from browsing items to checkout [3]. This aspect is known to be the most challenging for e-commerce businesses because high web traffic does not always equate to high conversion rates.

Market trends and customer behavior are aspects that CFS aims to predict as staying ahead of the trend allows for efficient product distribution and engaging marketing campaigns [3]. Authors [4] suggested that the problem with most predictions were that they were either self-serving to a company, unable to specify a clear time horizon, entirely prophetic without supporting evidence, or ignored external factors such as customer adoption or economic state.

2.1. Data Analytics Solutions

With the emergence of the Industrial Revolution 4.0, more companies have sought for solutions that involve data analytics to overcome these problems. This is because companies recognize that their daily operations generate large volumes of data, but they might not have enough time, money, or expertise to manually derive insights from the raw data. One of the most common types of data analytics in businesses is predictive analytics [5]. Predictive analytics analyzes past data to make future predictions using statistics, data mining, machine learning, and artificial intelligence [6]. It not only allows companies to learn what happened and why, but it also provides future insights. Predictive modelling involves six steps, namely project definition, source data exploration, data preparation, model construction, model deployment, and model management [7].

One vital component involved in the process of predictive analytics is using logistic regression to predict customer retention. Logistic regression is used to model the relationship between a dependent variable and one or more independent variables through calculating probabilities in a logistic equation [7]. Every independent variable has the coefficient 'b' that has an outcome of either 0 or 1, based on its contribution to changes in the dependent. There are binary, ordinal, and nominal logistic regression models. The formula of logistic regression is as seen in (1):

$$p = \frac{1}{1 + e^{-(b_0 + b_1x)}} \quad (1)$$

The p value represents the significance of the coefficient. The constant (b_0) navigates the curve left and right while the slope (b_1) determines the curve's steepness [8].

Authors [9] implemented a customer-based predictive analytics and optimization system for pricing decisions in a field experiment, combining store level panel data and pricing data from rival stores to form a demand model that had profitability as its dependent variable, as well as pricing and competition for the independent variable. They formed a predictive econometric model that estimated demand for a given item using price, feature, and display as variables. Market share was also modeled using time trends, seasonality, demand shifters, and special events to show effects of time on the estimation process. Each week, the optimized prices were tested in stores, the econometric model algorithm (Fig. 1) was re-estimated, and prices were re-optimized for the next week. In 13 weeks, their results showed that there was an increase in the gross profit margin by 40.7 cents per dollar [9]. The extrapolated total margin improvement per year for the enterprise was \$7,830,680 USD.

Aside from predictive analytics, web analytics are also used in companies that have an online presence or want to build one. It involves measuring, collecting, analyzing, and reporting on website data to understand and optimize website usage [10]. The most popular web analytics tool is Google Analytics as it is free but reliable. It provides results on offsite channels, applications, and websites. It uses first party cookies by Javascript to track user data such as behavior, demographic, and campaign data. It can track up to 50 profiles per account and 5 million page views each month. In addition, it can detect if a website is underperforming in terms of data security, web traffic or search engine optimization, so companies can make necessary improvements [10].

2.2. Problem Statement

CFS encountered two of the problems in section 2, namely inability to predict customer behavior and inefficient customer sales cycle tracking. The company failed to predict customer behavior when they launched their first collection [1]. The business had a customer base so big that the launch caused the website to crash and was nonfunctional for eight hours. Based on prior estimations, the website was primed to accommodate a few thousand visitors, but web traffic on the actual day overloaded the system [1]. On further inspection of their official Facebook page, they have encountered crashes or glitches for two launches in 2010, one in 2011, 2015, and 2016 each. It is safe to say that this is a recurring issue for the firm. This can cause a decrease in revenue for the firm as potential customers will leave to purchase items from websites with better web traffic accommodation.

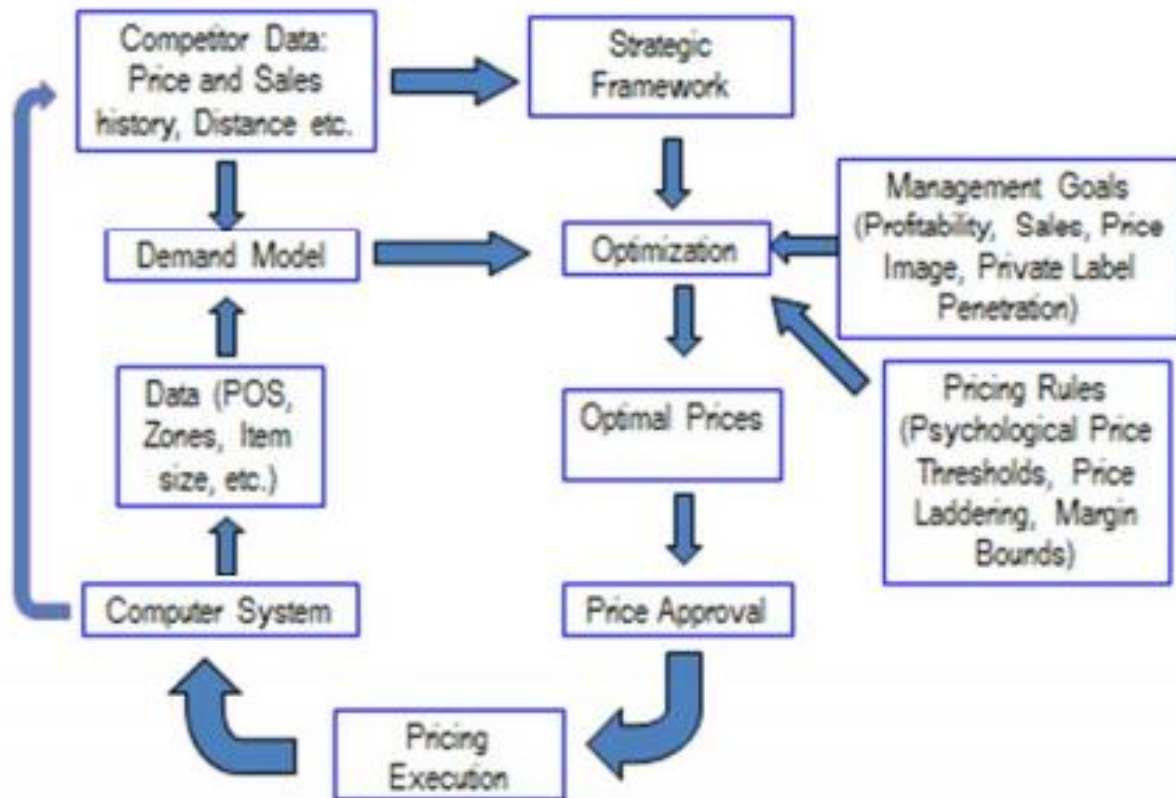


Figure 1. Algorithm flowchart for customer-centric predictive analytics [9].

CFS encountered inefficient customer sales cycle tracking in 2010. Competition among blogshops was tough, so even though their web traffic was high, sales conversion was low because users would browse the website without purchasing anything. Users could find similar items at cheaper prices in newer blogshops, so CFS lost its ability to stand out in the retail market [1]. Moreover, they were not tapping into the market that shopped in physical stores exclusively, so sales were not maximized.

2.3. Emergence of Data Analytics

CFS has a need for data analytics because the problems that they have encountered are problems that can be solved by analyzing data. Their daily transactions produce a large amount of data, so there is no shortage of data to be processed. Data analytics allow companies to accommodate high web traffic on their website through tools such as Google Analytics. It uses various metrics, namely new users, number of sessions per user, bounce rate, pageviews, page depth, time on page, and scroll depth to measure web traffic [11]. For CFS, the most important metrics to measure web traffic are new users and time on page. These metrics are important because the occasions that CFS has crashes are mostly new launches. New launches are not only explored by regular customers, but also by people who are just curious, and might not be looking to buy anything new for the time being. Therefore, the website traffic is expected to be higher than usual as a result of new users.

New launches also incur an increased average time on page. This is because many new items are released simultaneously, so users will be browsing several pages within one session to view the full catalogue of clothes. While this is good for increasing user engagement, bandwidth requirements also increase, causing websites to crash if not accommodated properly. From August 2020 to January 2021, CFS has the highest website visits at 520,000 in November 2020. This is mainly due to the 11.11 Singles Day celebration, where e-commerce firms have big discounts. CFS participates in this event, explaining the jump in visits. Google Analytics can be used to monitor website traffic during these peak periods, and the company can make the necessary adjustments to their bandwidth and server capabilities.

To increase customer retention and conversion, the company can also use Google Analytics as it provides valuable insights on customer behavior and engagement. This can be tracked using bounce rate and page depth as

metrics. Authors [12] defines a bounce as a session where the user opens a page on the site and exits without triggering further requests to the server for that page. CFS has an average bounce rate of 47.05%, which is common but can be decreased. Identifying the reasons behind the bounce points and making modifications will contribute to increasing customer engagement it may increase user page depth, thus increasing the likelihood of completing the transaction.

2.4. *Common problems faced by professionals*

As more businesses realize the importance of data analytics, professionals are often hired to assist them in managing various types of data. Two common problems faced by professionals in attempting to handle these data are synchronization and storage issues [13]. Synchronization of data is the process of establishing consistency of data over time between different data sources and a common repository [13]. This process is one of the most challenging as the tsunami of data that enters the repository from each source differs in time intervals and velocity. This might lead to inconsistencies or invalid data because the sequence of data does not align. When used in data mining, this data will lead to incorrect results. The heterogenous data also leads to difficulties in transforming and cleaning data for the ETL process. Fragmented data ownership contributes to data synchronization issues [14]. When data is owned by various stakeholders in different warehouses, the management of data into one consolidated warehouse is much harder as professionals need to go through various parties to obtain data that they need. These data might not be standardized, leading to difficulties in synchronization.

Data storage is another problem that is constantly faced by data professionals. Data professionals aim to derive information from data in order to make decisions. The more information the company has, the better the decisions made. Therefore, unstructured big data is required in large amounts. The problem arises when companies lack the devices required to store big data, leading to compromises on decision making [14]. Most systems can store up to 4 terabytes per disk. Big data is stored in exabytes, so 25,000 disks are required to store 1 exabyte. It is physically impossible for a system to contain that many disks [13]. Cloud computing is a potential solution for this issue, but transferring various data sources to and from the cloud is a time-consuming process [13]. Selective transfer of data for analysis can solve the problem, but it is difficult for companies to pick which data to store or skip. Therefore, more technology is required to discern useful data from the rest [14].

2.5. *Pros and Cons of Data Analytics*

CFS faces various advantages and disadvantages when using data analytics for decision making. There are two advantages of using data analytics, namely better satisfaction of customers' needs, and increasing efficiency of business operations. The usage of data analytics tools on websites such as Google Analytics allows CFS to understand their customer demographic because it inserts cookies into the website that records user attributes and activities on the website [15]. This allows the business to know the age, gender, and interests of customers. Google Analytics then segregates the data sent to the server by user, session, pageview, and event level [15]. The data collected can be further categorized into user acquisition data and behavior data. The former is data about users before they visit the website, while the latter is about users when they visit the website. User acquisition data allows businesses to conduct targeted advertisements to the demographic that is most likely to purchase their products, leading to greater success compared to generalized advertisements. This data also allows the company to provide customers with a user experience that is highly customized, increasing the likelihood of a customer completing the transaction. User behavior data identifies common points where users end their browsing session, allowing businesses to smooth out these issues to increase customer conversion.

The implementation of targeted advertisements via Google Analytics allows companies to capture leads early by increasing brand awareness among specific demographics [16]. For example, female fresh graduates should be targeted because they will be looking for office wear. By introducing CFS to customers before they even think of buying clothes, this builds familiarity of the brand with users. This subtly persuades customers into choosing their products instead of rival brands when completing the final purchase. Therefore, this increases the efficiency of business operations.

The main disadvantage of using data analytics for decision making is that data analytics requires high costs for implementation [17]. Implementation of data analytics typically involves three components, namely infrastructure, maintenance, and human resource. Infrastructure requires an upfront payment for materials such as data storage, servers, network, and monitoring tools. Moreover, designing, coding, testing, and documenting has to be done. The total cost of these requirements is proportional to the data platform's size [17]. To illustrate, the estimated price of a 3 terabyte per month database is more than \$134,000 USD [18]. After the infrastructure is set up, maintenance and manpower are then required to sustain the overall data analytics lifecycle [17]. Data professionals will be hired to

facilitate future development and improvement of the data platform, further consuming time and resources. These combined factors are the reason why some businesses are not willing to invest in data analytics.

3. Proposal of Data Analytics Solution

As seen in subsection 2.5, the advantages of implementing data analytics outweigh the disadvantages. Thus, two data analytics architecture models are proposed for CFS to overcome their business problems.

3.1. Website and Network Monitoring

The implementation of Google Analytics will help CFS to better predict customer behavior on the website. This is because it collects user behavior data when they browse the website [15]. These include the duration of browsing, first and last page on the website, as well as pathway into the website. Based on user behavior data, web developers and engineers can analyze where customers stop browsing in order to fix these issues. Ultimately, Google Analytics reports act as a guide for the company to make decisions for the website, providing a better user experience as it is backed by genuine user data [15]. To better manage high web traffic on CFS website, Firebase Crashlytics must be implemented in CFS website on top of Google Analytics. This is because Google Analytics is only capable of analyzing user data on the website, but not bandwidth and bugs that lead to crashes. Firebase Crashlytics is a Google developed crash reporting solution for websites and applications [19]. It helps websites by tracking, prioritizing, and fixing stability issues that affect website quality. It also decreases the time required for troubleshooting through categorizing crashes and highlighting the causes. These insights let website engineers know if a crash has impacted many users, and which lines of code are causing them. Realtime alerts are also sent when an issue becomes more serious. The best feature of Firebase Crashlytics is that it provides custom crash reports for each incident, providing a list of issues in terms of severity and prevalence to help engineers fix them quicker.

Since CFS website integrates third-party APIs for e-payment transactions and advanced authentication, Firebase is positioned between CFS server and clients [20]. This allows the server to connect to Firebase Crashlytics and access client data. The server code observes any changes to client data, and responds accordingly. In this architecture (Figure 2), the Firebase handles all of the major issues in terms of web traffic instead of the server. This allows for better server function and efficiency, as Firebase monitors the lack of bandwidth and the presence of bugs via client data before it is sent to the data source.

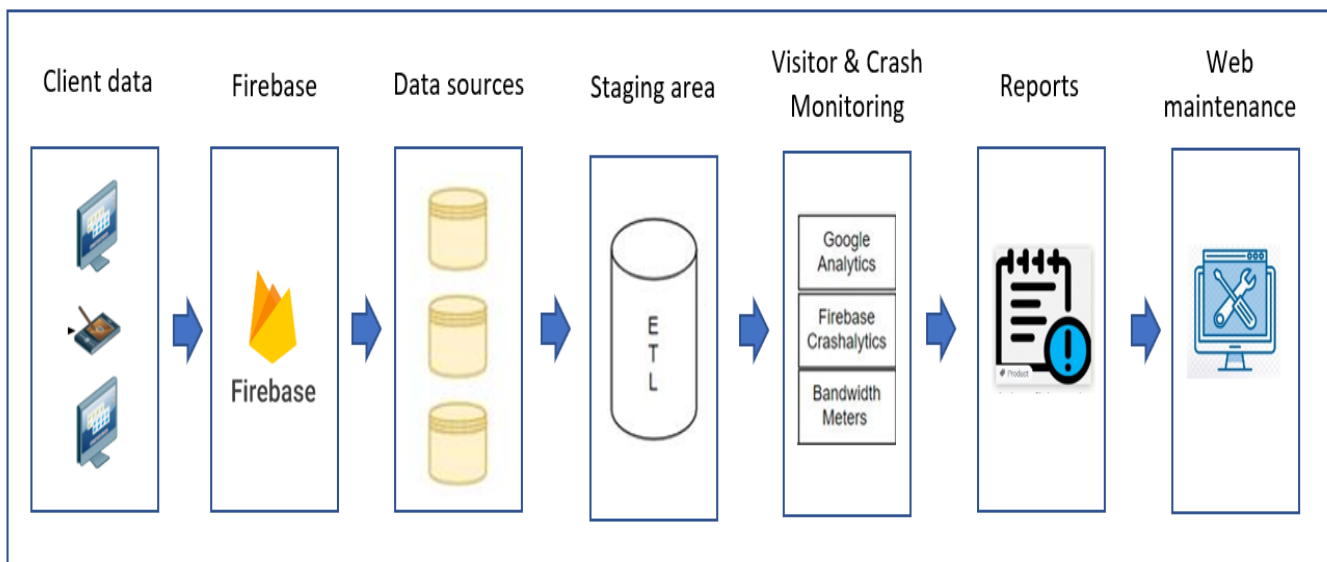


Figure 2. Architecture for Website and Network Monitoring.

When client data is obtained from user devices, it is processed through Firebase which scans for crashes and errors, providing real time updates. Then, client data is stored in data warehouses, allowing the company to prepare it for the extract, transform, and load (ETL) process. After ETL, the next step analyzes information derived from Google Analytics, Firebase Crashlytics, and bandwidth meters. Bandwidth meters are used to support the first two tools because they monitor the performance of the network as a whole, including WAN or LAN utilization [21]. It allows IT professionals to set traffic and device limits that highlight network activity and inform them when

those values have been exceeded. Better understanding of bandwidth utilization will ultimately lead to better capacity planning and scaling. Therefore, performance bottlenecks and connectivity issues faced by users can be overcome and improved [21].

The insights generated from Google Analytics, Firebase Crashlytics, and bandwidth meters will then be generated into reports used to make adjustments to either the website or network settings. The reports will be the main guide to the last step. Web maintenance requires web engineers to execute all required updates to the network system or website function. The aim of this architecture is to provide the best website experience. It is suggested to be executed regularly (before each new launch) in order to maintain consistency for users.

3.2. Customer Conversion and Retention

To increase CFS customer conversion rates, targeted advertisements should be implemented regularly via various platforms such as Google Advertisements, Facebook and Instagram to narrow down the potential customer pool. Targeted advertising involves online advertising which is directed towards audiences with specific traits that are related to the product the advertiser is promoting [22]. These traits can be categorized into demographic, psychographic, or behavioral traits [22]. Targeted advertising relies on these traits to ensure that the right advertisements reach the right people. The best method to execute targeted advertisements for CFS is to use a decision tree model (Figure 3).

4. Results and Discussion

Decision trees use a dichotomous tree constructed from nodes, and subdivides the population into groups based on variables [23]. The best possible division is selected until further separation is impossible. This method is the best for targeted advertising because it gives unambiguous results and is not affected by missing data. CFS model involves variables such as sex, age, employment status, and socioeconomic status (SES). It uses IF THEN ELSE statements to evaluate future actions. One thing to note about this model is that these variables do not take into account other factors such as keyword searches that might affect the advertising process.

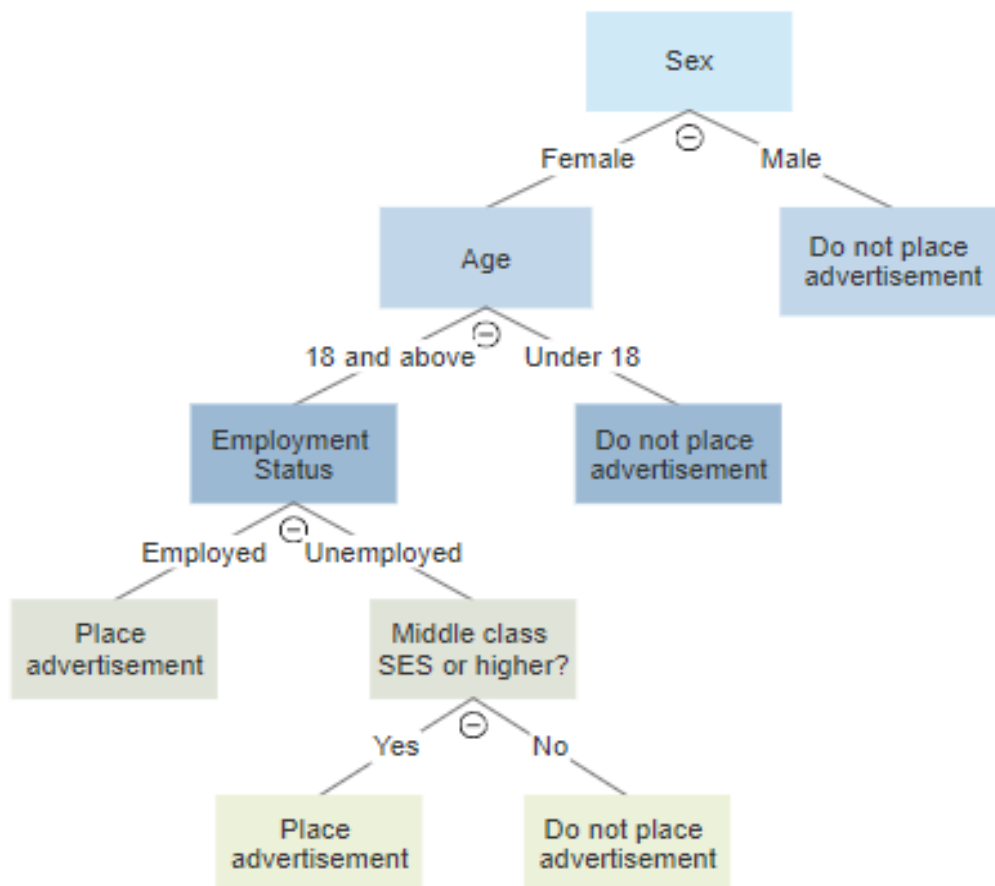


Figure 3. Decision Tree Model for Targeted Advertising.

Next, advertising data is stored in the company's data warehouse. After the advertisements have been displayed, engagement must be recorded and analyzed in reports. Metrics such as conversions, cost per conversion, leads, and cost per lead are used in this process. These insights can be obtained from the advertising platforms, namely Google Analytics, Facebook Audience Insights, and Instagram Insights. A baseline KPI is established to compare future performances for growth, and goals are set regularly to improve the baseline. These reports will act as guides for web engineers when executing future advertisements and making changes to the website. This establishes an integrated system with data marts where web analysts and engineers can pull data for any ad hoc decisions (Figure 4).

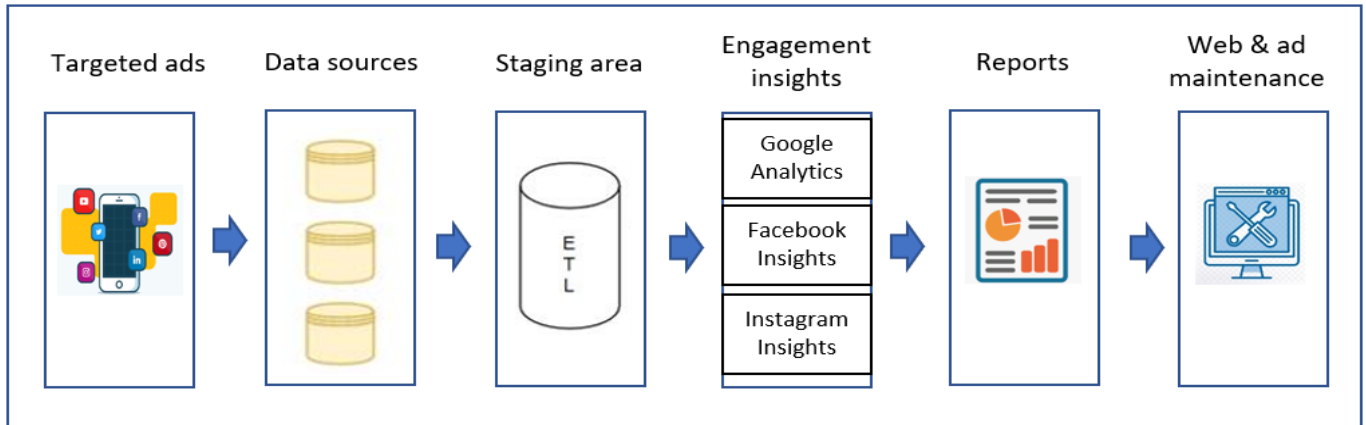


Figure 4. Architecture for Customer Conversion and Retention.

In both architectures, the main challenge in its development and implementation is user data privacy [24]. Privacy issues include controlling how companies access, modify, and utilize personal data. Individuals must be consenting to the utilization of their data and the types of data that can be accessed. They must be able to modify data about themselves. Lastly, they need to trust that companies will fulfil their duties and use the data for its sole purpose, ensuring its security [24].

5. Conclusion

Moreover, data warehouses in terms of architectural built of its integrated data is trustworthy and should be implemented by enterprises, businesses, and retails. It delivers high-tech business intelligence, it saves a jack pile of time, consistency and data quality can be ensured. It also generates a soaring return of investment although in the initial stages, does cost a big check. Decision making has never been easier with the help of a data warehouse and data integration. Businesses and retail services will skyrocket - all thanks to data integration and data warehouses. Thus, effective storage management and data integration are needed for better retail system purposes. When the system is integrated, there are still many issues to be solved in the retail SME area, such as data security, privacy of sensitive customers' data, and retrieval of information as such customers can access their personal buying records from any branches or platforms and their wearable device through app. In addition, small and medium-sized enterprises and clinics have limited resources in human power and finance to digitize their paper written records. Transforming from a physical record to an electronic record system requires a large amount of training and adaptation to the staff. Hence, efforts from the private and public sectors are required establishing a more advanced, efficient information exchange system.

BIBLIOGRAPHY

- [1]. Z. Salim, C. Ooi, W. Yan, M. Petraeus, and Y. Sivakumar, "Built on just \$500, this s'pore brand has now grown into a multi-million fashion powerhouse," Vulcan Post, 30-Jun-2021. [Online]. Available: <https://vulcanpost.com/628312/love-bonito-turned-500-into-multi-million-fashion-powerhouse/>. [Accessed: 12-Mar-2022].

- [2]. “How data analytics impacts small businesses in 2023,” business.com. [Online]. Available: <https://www.business.com/articles/the-state-of-data-analytics-in-2019/#how-can-smbs-use-data-analytics>. [Accessed: 13-Mar-2022].
- [3]. R. Rahutomo, Y. Lie, A. S. Perbangsa, and B. Pardamean, “Improving conversion rates for fashion e-commerce with a/B testing,” 2020 International Conference on Information Management and Technology (ICIMTech), 2020.
- [4]. K. Lobaugh, B. Stephens, C. Reynolds, R. Bamford, and S. Davis, “The future is coming ... but still one day at a time,” Deloitte Insights. [Online]. Available: <https://www2.deloitte.com/us/en/insights/industry/retail-distribution/retail-and-consumer-products-predictions.html>. [Accessed: 30-May-2022].
- [5]. H. Surbakti and A. Ta’a, “Managing knowledge business intelligence: A cognitive analytic approach,” AIP Conference Proceedings, 2017.
- [6]. V. Kumar and M. L., “Predictive analytics: A review of trends and Techniques,” International Journal of Computer Applications, vol. 182, no. 1, pp. 31–37, 2018.
- [7]. O. K., O. O., and A. O., “Predictive analytics for increased loyalty and customer retention in telecommunication industry,” International Journal of Computer Applications, vol. 179, no. 32, pp. 43–47, 2018.
- [8]. K. Shah, H. Patel, D. Sanghvi, and M. Shah, “A comparative analysis of logistic regression, Random Forest and KNN models for the text classification,” Augmented Human Research, vol. 5, no. 1, 2020.
- [9]. E. T. Bradlow, M. Gangwar, P. Kopalle, and S. Voleti, “The role of Big Data and predictive analytics in retailing,” Journal of Retailing, vol. 93, no. 1, pp. 79–95, 2017.
- [10]. V. Kumar and G. A. Ogunmola, “Web analytics for knowledge creation,” International Journal of Cyber Behavior, Psychology and Learning, vol. 10, no. 1, pp. 1–14, 2020.
- [11]. T. Semerádová and P. Weinlich, “Using google analytics to examine the website traffic,” Website Quality and Shopping Behavior, pp. 91–112, 2020.
- [12]. D. Walsh, P. Clough, M. M. Hall, F. Hopfgartner, J. Foster, and G. Kontonatsios, “Analysis of transaction logs from National Museums Liverpool,” Digital Libraries for Open Knowledge, pp. 84–98, 2019.
- [13]. M. A. Wani and S. Jabin, “Big data: Issues, challenges, and techniques in business intelligence,” Advances in Intelligent Systems and Computing, pp. 613–628, 2017.
- [14]. H. Surbakti and A. Ta’a, “Tacit knowledge for business intelligence framework: A part of unstructured data?,” Journal of Theoretical and Applied Information Technology, vol. 96, no. 3, pp. 616–625, Feb. 2018.
- [15]. M. Bala, D. Verma, “A Critical Review of Digital Marketing”. International Journal of Management, IT & Engineering, 8(10), 321–339, 2018.
- [16]. A. Brooks, “7 ways a mere exposure effect can boost conversions,” Venture Harbour, 03-Nov-2021. [Online]. Available: <https://www.ventureharbour.com/the-mere-exposure-effect-7-ways-to-use-it-to-boost-your-conversions-with-examples/>. [Accessed: 17-May-2022].
- [17]. Y. Wang, L. A. Kung, and T. A. Byrd, “Big Data Analytics: Understanding its capabilities and potential benefits for healthcare organizations,” Technological Forecasting and Social Change, vol. 126, pp. 3–13, 2018.
- [18]. [18]. J. L. Leevy, T. M. Khoshgoftaar, R. A. Bauder, and N. Seliya, “A survey on addressing high-class imbalance in Big Data,” Journal of Big Data, vol. 5, no. 1, 2018.
- [19]. “Firebase Crashlytics,” Google. [Online]. Available: <https://firebase.google.com/docs/crashlytics>. [Accessed: 12-May-2022].
- [20]. J. García, V. De Lera, D. Lacambra, F. Gimeno, and Á. Alesanco, “Design and development of an Android app based on Firebase in the study of unsportsmanlike behaviors in grassroots tennis competitions,” Journal of Mobile Multimedia, 2021.
- [21]. T. T. Oo, “Design and implementation of bandwidth monitoring, line aggregation of VoIP,” International Journal of Advanced Trends in Computer Science and Engineering, pp. 1326–1331, 2019.
- [22]. J. Shin and J. Yu, “Targeted advertising and consumer inference,” Marketing Science, vol. 40, no. 5, pp. 900–922, 2021.

- [23] . H. Lipyanina, A. Sachenko, T. Lendyuk, S. Nadvynychny, and S. Grodskyi, “Decision tree based targeting model of customer interaction with Business Page,” *Computer Modeling and Intelligent Systems*, vol. 2608, pp. 1001–1012, 2020.
- [24] . I. Someh, M. Davern, C. F. Breidbach, and G. Shanks, “Ethical issues in Big Data Analytics: A stakeholder perspective,” *Communications of the Association for Information Systems*, pp. 718–747, 2019.