**Research Project**

William Toland

22FD-MIS581-1 Capstone: Business Intelligence and Data Analytics

CSU Global

Dr. Lisa Lee Bryan

November 27th, 2022

**Abstract**

Ecommerce has grown significantly within the past decade. Businesses are being required to implement web ordering platforms at an exponential rate to stay competitive within their given market. This research paper aims to examine several key aspects of Motion Industries online web ordering platform. The three questions answered within this research are as follows. First, how can Motion improve ecommerce customer churn? Second, how can Motion improve ecommerce sales through analysis? Finally, how can analysis predict future ecommerce sales?

The hypothesis testing within this research paper utilizes several key tools. These tools being Python, SAS OnDemand, Qlik, Excel, and SQL coding. The data generated comes from an established Qlik dashboard that provides customer data based on customers online web activities. Key variables used include monthly sales data, product impression data, and cart additions.

The analysis showed that ecommerce customers that had online sign in activity of less than 15 days between sign in attempts spent significantly more on the ecommerce platform. Analysis also revealed a moderate correlation between cart additions and product detailed impressions. Finally, analysis showed that the monthly web sales is in a strong uptrend position with an average monthly growth percentage rate of 24%.

Each of these hypotheses help answer each of this research projects questions. Customer churn can be improved by improving customer website activity and average days between sign in attempts. Motion can improve ecommerce sales by improving website functions that result in more detailed product impressions. Finally, this research helps show how Python can be used to indicate sales trends to help predict future sales.

**Introduction**

Ecommerce analytics focuses on finding and communicating shopping and web-based patterns with business decision makers to improve ecommerce functions. Ecommerce analytics uses an accumulation of data from areas of the website that can be then used to understand what, when, how, and why customers are purchasing items on the site. The insights driven by ecommerce analytics can help businesses improve the way their items show up, what items are recommended, how prices and inventories are playing a role, and what customers are purchasing. The main aspects of ecommerce analytics are based on the audience, acquisition, behavior, conversions, and marketing. This research project focuses on the ecommerce website provided by Motion Industries. Motion Industries is an industrial parts distributor for various products including bearings, transmissions, electrical, hydraulic, etc. The organization works throughout all North America. The aim will be to analyze consumer behavior through various means of analysis. The datasets used in this research will focus on customer web sales, internal search history, customer login data, and various other analytics to create forecasts, trends, and insights.

**Objectives**

As time continues the use of data has become an increasingly large component of business performance. According to Akter & Wamba (2016), productivity is increased by 5-6% when e-commerce firms help implement big data analytics. Akter & Wamba (2016), also mention that big data analytics also contribute 10% or more of the growth for 56% of firms.

To easily analyze big data from an enterprise level, it is important to establish an enterprise customer relationship management system which allows organizations a way to respond to data efficiently and effectively. However, the issue with CRM systems is that businesses need to answer four main questions. First, is the CRM strategically solving all of Motion’s organizational needs? Where are Motion’s issues regarding the companies CRM platform? Does Motion need to gather better data? Where does Motion go from here in relation to the company’s data (Ledingham & Rigby, 2004)? These four questions must not only be answered but monitored as organizations continue to improve and develop.

The issue with Motion Industries ecommerce analytics is that the organization does not have a means to analyze data. Motion uses an Enterprise CRM platform to analyze large amounts of data including over 256,000 products, 100,000s of customers, locations across North America, and as of the past couple years web sales data. As Motion has grown the organization has found a large need to establish an online web presence. The organization implemented a new website and began tracking and improving the website based on several forms of data. The issue, however, is that the organization is not currently monitoring consumer behaviors. While the organization is collecting the data, it has no applications that monitor it to give teams around Motion the tools, they need to improve web functions.

Motion gathers web user data including sales amount, order amount, cart additions, internal search phrases, types of items purchased, company names, and web user login ids. Even with the data collected, there are no applications in place that analyze said data to provide insights into the customers web purchasing activities. By understanding the data being brought in by the website, Motion gains several advantages. It would allow the organization to reach a wider audience. By understanding which organizations are purchasing from the web platform and what they are purchasing, it allows Motion to expand product categories. Motion can also work to improve business from organizations not purchasing. This could include lowering product price offerings or increase product lines. Another key factor is maintaining web user activity and sales trends. Without a system to accurately monitor this, there is no way to keep customer purchasing or following up with customers that are trending downwards regarding web activity.

Monitoring the website will allow Motion a way to use data-drive insights. The organization can use an app to help gain inferences on what is potentially not working. Decision by executive members can be done to improve functions and make the online shopping experience much better. Taking guess work out of website optimization leads to long term online success and presence in the distribution industry.

**Benefits**

Once an application with KPI’s and visualizations is implemented Motion will gain several benefits. Website experience managers will be able to accurately monitor consumer behaviors, helping them create strategic plans to improve consumer experiences. It will also allow them to improve web functions based on company wide web trends. If certain items are being added as cart additions, however not ordered, it allows the company to improve the margins on these items or dive deeper into why they are not being ordered. An application can also be implemented to sales representatives to understand how their consumer is using the website and allows for an opportunity to reach out to the customers to gain more sales.

**Overview of Study**

The aim of my research is to create an application that utilizes several KPI’s to accurately provide Motion with a means to monitor customer behavior. This will be done mainly via Qlik which utilizes SQL and in app tools to create tables and visualizations. The first step will be to establish a table that consists of drill down features using customer information. The table will also include critical KPI’s and metrics that can be used to gain insight into customer behavior. Metrics will include like previously mentioned customer close rates, cart abandonment rates, cart addition to order rates, average orders per sale, average time between logins, manufacturing data, etc. Once the metrics are in place the next goal will be to establish baseline measurements using SAS, R Studio, or Python. While it is important to have these metrics, it is more important to be able to understand how to use them. Finally, various visualizations will be established to help show trends.

Visualizations can monitor trends such as certain item categories trending upwards. It can also be used through a drill down option to find individual products trending upwards. The data can also be used on a consumer level to visualize organizations that have declining sales figures. This data can then be used by sales or experience teams as an alarm to reach out to customers.

**Research Questions and Hypothesis**

To improve Motion’s ecommerce business functions, it is critical that several questions are answered. First, how can Motion improve ecommerce customer churn? Second, how can Motion improve ecommerce sales through analysis? Finally, how can analysis predict future ecommerce sales? Each of these questions will be answered through hypothesis testing as shown below.

**Hypothesis 1**

Ha: Ecommerce customers that go over 15 days of inactivity result in lower sales.

H0: Ecommerce customers that go over 15 days of inactivity result in zero or equal sales amounts.

**Hypothesis 2**

Ha: Ecommerce sales has a relationship with detailed impressions and cart additions.

H0: Ecommerce sales exhibits no relationship with detailed impressions and cart additions.

**Hypothesis 3**

Ha: Ecommerce sales show positive monthly trending sales.

H0: Ecommerce sales show negative monthly trending sales.

Each of the above hypothesis will be answered through statistical testing. Hypothesis one will be answered using Qlik by leveraging data based on average sign in and sales activity. The averages will be compared to find sign in activity averages alongside customer sales data. If the average website sales increase for customers with sign in activity less than 15 days, then we would reject the null hypothesis. The second hypothesis will be analyzed using SAS OnDemand via correlation analysis and modeling. The correlation coefficient will be used to test the hypothesis with a score of +1 indicating that a perfect linear relationship exists between sales, detailed impressions, and cart additions (Mukaka, 2021). If cart additions, sales, and detailed impressions show a strong correlation then the null hypothesis would be rejected. Finally, hypothesis three will be done within R Studio via regression analysis. The regression analysis will help determine if monthly sales over a 12-month period are trending positively or negatively. If monthly sales show a positive trend upwards, then the null hypothesis would be rejected.

**Literature Review**

**Ecommerce Analytic Approaches**

There are many techniques and tools that organizations can leverage to improve e-commerce platforms. When analyzing large data, it is advisable to focus on small data sets which are based on which large data sets can be estimated and predicted (Kumar & Chandrasekar, 2016). The techniques that can be used to analyze data under these situations can be Apriori algorithms, probabilistic approaches, clustering, and regression. Apriori algorithm is an algorithm for mining associated that was developed in 1994 (Haoyu, 2021). This algorithm identifies frequent individual items in a database and extends them to larger item sets if item sets appear often in the database. The most common example of the use of the Apriori algorithm is its use in market basket analysis (Haoyu, 2021). The probabilistic approach assumes that randomness plays a large role in predicting future events. The probabilistic model gives a probability distribution as a solution through models such as the Naïve Bayes method (Alon et al., 2000). Regression analysis allows the examination of the relationship between two or more variables of interest. These various techniques all play a large role when analyze e-commerce data.

Kumar & Chandrasekar’s (2016) study also offers several forms of analytics in relation to e-commerce data. First, text and web analytics which are used to extract information from unstructured text to create structured data patterns (Miner et al., 2012). This type of analytics is particularly helpful when analyzing ecommerce consumer website search results. Analyzing unstructured text phrases may help find trends that allow for organizations to improve strategic e-commerce marketing approaches. Next, the use of mobile analytics is suggested because it goes hand in hand with e-commerce. Monitoring m-commerce allows organizations to gain valuable data about customers. For example, monitoring how often the customer is ordering through their mobile shows that this customer is constantly busy. In the manufacturing world through inferences this could indicate that the customer is working in multiple roles and the business being called upon is smaller in nature.

**Research Design**

The type of research being used within this study focuses on a quantitative approach. The quantitative approach according to Albers (2017), utilizes numerical data that must be analyzed to assist drawing study conclusions. This methodology utilizes critical thinking in relation to how data is analyzed. The data used within my paper as shown in Figure 1 uses consumer data based around website sales. The data has been collected through customer web sales activity. The variables collected include item numbers, sales amounts, number of cart additions, customer id’s, etc.

**Methodological Connection**

Through quantitative research both insights and tools will be established to analyze online web users’ data. Utilizing the quantitative approach will allow the creation of both key metrics through analyzation and critical visualizations. The goal of my research is to provide enough insight into online customer behavior that it enables Motion experience managers the ability to improve future web sales.

**Instruments and Dataset**

The dataset that will be used within my project will be based on data gathered from my organization’s website. The data will be collected based on customer website use and order details. The data will include several excel sheets consisting of various variables. Variables that will be analyzed include customer level information, organizational location information, sales information, and manufacturer data (Pérez-Rodríguez et al., 2013). The data will come from the main tool my organization utilizes which is Qlik. All the data has been set up within databases and connected to Qlik to offer easy access to the data that will be needed for my research project. Each type of data will be able to be extracted into excel format to offer analyzation in bother Python and SAS OnDemand.

**Tool and Techniques**

There will be three tools being used within this project. First, Qlik will be used to establish a dashboard or application that business experience management can use to gather quick insights into how customers are using our organizations website to order items. This dashboard will be utilizing SQL to input the data needed from the database into the Qlik platform. Several options will then be used to create an application. These options include the use of drill down features and filter panes to break down several key variables. Next, several important KPI’s will be established utilizing Qlik code and basic math equations. Gräve (2019), states that the most important KPI’s when looking at ecommerce metrics are cart attainment rates, average sales amount per order, and cart addition data. Each of these KPI’s will be included within my Qlik application alongside graphs that assist in visualizing the month over month performance of each KPI.

Next, python will be used to further break down the dataset’s variables. Several techniques will be used such as lifetime value prediction, churn analysis, and sentiment analysis. Data exploration will also be done within Python to potentially uncover valuable insights. The last use of Python will be based around finding out the best value rate that leads into sales. This will be used to better understand KPI’s such as product views per order. It will allow for significantly less data inference (Taylor & Todd, 1995).

Finally, SAS OnDemand will be used to provide predictive analytics on sales data. Linear models will be used for 15 of the customers alongside the data at a territory level. Clustering models will also be created based on manufacturing sales data to uncover items being sold together.

**Data Type Variables**

There are several types of data variables being used within this research project. Organization branch names, customer names, manufacturer names, product categories all represent categorical data. Sales figures, number of orders, number of cart additions, and product quantities all represent continuous numeric variables (Gräve, 2019). Attainment rates represents a discrete numeric variable. The orders data will be created from binary Boolean factors. If a customer places an order, then it is represented as 1 or true and then inputted into the dataset. Period variables will also be utilized to forecast this data will include continuous data over a 12-month span.

**Figure 1**

*Data dictionary*

*Table

Description automatically generated with medium confidence*

Note. Figure 1 shows the intial data dictionary for dataset being used for research project.

**Data Potential**

The data within my dataset plays a large role in how my organization handles their website and consumers. Alrumiah and Hadwan (2021), state that many benefits come from organizations having the ability to utilize ecommerce analytics. It offers organizations the ability to further understand consumer behaviors. This allows organizations to increase sales based on website improvements. It also allows organizations to personalize website recommendations based on what is being ordered and looked at. Furthermore, it allows organizations to improve marketing efforts based on items being sold. Finally, it allows pricing and inventory to cater to the needs of consumers. By analyzing this dataset, I will provide strategic information into how consumers purchase and will offer techniques that can be used by my organization to forecast sales of various products. The created application in Qlik will provide web sales experience managers extra tools to help improve both the website and how customer are utilizing the website.

The Qlik application will provide information into how many items are being purchased from a month-to-month basis. The information can then be used to improve consumer relations based on if sales are declining. It can also be used by the organization to uncover why customer sales are improving and provide insights into what can be better managed to help improve other website customers.

**Limitations**

Within my study there are a few limitations to consider. Limitations defined by Ross and Zaidi (2019) are weaknesses within a research design that can influence the outcomes and conclusions of the research. The largest limitation within my study focuses on the amount of historical data that is being used within this research project. The web sales data has been collected within the past two years resulting in only two years of historical data. While historical data can be used for analyzation the more data that is available the more accurate the models will be (Sutton & Austin, 2015). The next limitation is based on timing. Motions website was established and built during the peak of COVID-19. The data collected during the pandemic could potentially skew the results of the project. This limitation also co-coincides with the lack of historical data as research cannot be measured pre and post pandemic.

**Ethical Considerations**

Another critical component within this research design focuses on ethical considerations based on the data being utilized for this project. Hand (2018), states that ethical issues involving data can be more challenging than ethical challenges based on other technologies. The ethical issues involving data can impact all aspects of an organization due to the data intrinsic value. David (2018), proposes a checklist of the following:

* Identify which ethics body has oversight of work;
* Be aware of institutional policies;
* Be aware of national regulations and laws;
* Keep record of data modifications and manipulations;
* Understand data’s origin;
* Treat meta data as rigorously as base data;
* Store data securely;
* Determine how long the data can be kept;
* Specify data access;
* Ensure the appropriate tools are being used to analyze data;
* Have systems in place that allow data to be corrected;
* Be clear about benefits of analysis and who benefits.

Each of these steps play an integral role within data analytics. The data that I am using is based around Motion’s customer level data. This data is based on individuals that work for various organizations. The roll up features allows for entire organizations to be analyzed. The data is stored on a work laptop that uses a VPN system alongside various security measures to ensure the data is safe. The data being used is also given based on special data access to ensure those that have access to the data can be monitored. Within my research the data being used for this research project focuses on one branch location with around 1,000 customers. If this paper were to be shared, it would be important that the entire organizations sales figures are not shared with competitors. Competitors could gain several advantages through the data being used within this research project. First, competitors could use it to offer more strategic pricing on competing products. Second, competitors could solicitate consumers they are not currently working with.

**Data Analysis**

**Qlik and Excel**

Data analysis was done using four separate tools. First, Qlik was used to create csv files for importation into Python, Excel, and SAS OnDemand. Hypothesis one testing required a table to be created based on average customer sign ins. This was done by importing customer sign in dates, then converting these dates into integers. Once this was done the days between sign ins were calculated and averaged. Once finished, the ability to select customers based on average sign in activity of less than 15 days was possible. The results in Figure 1, shows that sales increased dramatically if customers average sign ins were lower than 15 days. This indicates that the null hypothesis should be rejected.

**Figure 1**

*Average sign in data for total web sales*

**Table

Description automatically generated**

Note. Figure 1 showcases the average monthly sales beginning January 2022 to October 2022 for sign in users below average 15 days between logins and users over 15 days between sign in.

**SAS OnDemand**

Hypothesis two was tested using SAS OnDemand. This hypothesis used data pertaining to cart additions and detailed product views. Detailed product views was data collected based on customers clicking onto products and viewing them for x amount of time. To test hypothesis two, correlation analysis was done to find the variables Pearson Correlation Coefficient. The Pearson Correlation is used to access the linear association between cart additions and detailed product views. According to Mukaka (2012), a coefficient value lying between ± 0.30 and ± 0.49 indicates a moderate correlation between the variables.

The Pearson Correlation Coefficient shown in Figure 4 indicates a moderate correlation, with a value of 0.474 between cart additions and detailed product views. To take this a step further linear regression was done within SAS as shown in Figure 5. The linear regression resulted in a p-value which was less than .05 indicating a statically significant relationship between cart additions and detailed product impressions. Based on these models the second hypothesis H0 would be rejected.

**Figure 2**

*SAS OnDemand data*

**Graphical user interface, table

Description automatically generated**

Note. Figure 2 showcases the type of data uploaded to SAS for analysis.

**Figure 3**

*Scatterplot Matrix for detailed product impressions and cart additions*

**A picture containing graphical user interface

Description automatically generated**

Note. Figure 3 is a scatterplot which includes two of the variables being analyzed for hypothesis two.

**Figure 4**

*Correlation coefficient for cart additions and detailed product view*

**Table

Description automatically generated**

Note. Figure 4 shows the results for the correlation analysis done within SAS.

**Figure 5**

*Linear regression for cart additions and detailed product views*

**Table

Description automatically generated**

Note. Figure 5 shows the results of linear regression using cart additions as the dependent variable and detailed product views as the continuous variable.

The final hypothesis was tested through Python and Excel. This was done by looking at monthly sales data between January of 2021 and October of 2022. This was done by creating a linear chart to visualize the trend of sales. Figure 6 includes the results, this being a chart that shows a strong uptrend. Next, the table was rearranged within Python and exported to excel to calculate average percentage of change over the data’s timetable. This resulted in an average web sales growth of 24% per month. This average growth indicates a large uptrend of web sales from month-to-month. The result from this analysis indicates the H0 for hypothesis three should be rejected.

**Figure 6**

*Python monthly web sale plot*

**Chart, line chart

Description automatically generated**

Note. Plot created within Python represents monthly web sales between 2021-01 to 2022-09.

**Figure 7**

*Python data extracted to Excel*

**Table

Description automatically generated**

Note. Python data was exported to excel. Monthly web sales percentage of change calculated for each month then averaged.

**Conclusion**

Motion Industries has placed a large emphasis on improving the organizations ecommerce platform. This research paper helps showcase improvements alongside important findings that could benefit the organization. Several tools were used to help test several hypotheses. The tools used within this study were Qlik, Python, SAS OnDemannd, and Excel. The results of this study showed that the web site has had dramatic improvements and that the websites growth is on a steep incline. The study lists the importance of data based around customer sign in activity. Finally, this research shows correlation between key variables that may be key in improving web sales further.

**Recommendations**

The first recommendation, would focus on Motion implementing more analytics tools based on customer activity. This would largely drill down to how long customers are on the website, average products being searched in one session, and possibly a decision tree based on key phrases customers are searching. More data and analytics based on these key variables could help gather insights on the best customers for digital experience managers to target. The second recommendation would be based on more data based around product impression. The research indicated a correlation between impressions and cart details. However, the impressions are a blanket variable and could be further detailed to gain a better understanding of the correlation. Finally, growth percentages from the organization alongside drill-down growth for each location should continuously be monitored to help find trends that may improve the ecommerce platform.

**References**

Aivalis, CJ and AC Boucouvalas. (2011). Log file analysis of E-commerce systems in Rich Internet Web 2.0 applications. *2011 15th Panhellenic Conf. Informatics (PCI 2011)*, 30 September–2 October 2011, Kastonia, Greece, pp. 222–226. [Crossref](https://www.worldscientific.com/servlet/linkout?suffix=S2424862220500207BIB001&dbid=16&doi=10.1142%2FS2424862220500207&key=10.1109%2FPCI.2011.31) [Business Analytics in E-Commerce: A Literature Review](https://www.worldscientific.com/doi/abs/10.1142/S2424862220500207).

Albers, M. J. (2017). Quantitative Data Analysis—In the Graduate Curriculum. *Journal of Technical Writing and Communication*, *47*(2), 215– 233. <https://doi.org/10.1177/0047281617692067>

Alon, Noga; Spencer, Joel H. (2000). [*The probabilistic method*](https://amzn.to/3ml4kYI)*(2ed)*. New York: Wiley- Interscience. ISBN 0-471-37046-0.

Alrumiah, S., & Hadwan, M. (2021). Implementing big data analytics in e-commerce: Vendor and Customer View. *IEEE Access, 9*, 37281-37286.

Akter, S., Wamba, S.F. Big data analytics in E-commerce: a systematic review and agenda for future research*. Electron Markets 26*, 173–194 (2016). <https://doi.org/10.1007/s12525-> 016-0219-0

David J. Hand. Aspects of data ethics in a changing world: Where are we now?. *Big Data.*  Sep 2018.176-190.[http://doi.org/10.1089/big.2018.0083](https://doi.org/10.1089/big.2018.0083)

Gräve, J.-F. (2019). What KPIs Are Key? Evaluating performance metrics for social media influencers. *Social Media + Society, 5(3).* <https://doi.org/10.1177/2056305119865475>

Kumar, Prem & Chandrasekar, Sabarinathan. (2016). E-Commerce Trends and Future Analytics Tools. *Indian Journal of Science and Technology. 9.* 10.17485/ijst/2016/v9i32/98653.

Miner, G., Delen, D., Elder, J., Fast, A., Hill, T., & Nisbet, R. A. (Επιμ.). (2012). Chapter 16 - The Future of Text and Web Analytics. *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications* (991–1005). doi:10.1016/B978-0-12-386979- 1.00044-X

Mukaka M. M. (2012). Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi medical journal: the journal of Medical Association of Malawi*, *24(3),* 69–71.

Ning, J., Babich, V., Handley, J., and Keppo, J. (2018). Risk-aversion and B2B contracting under asymmetric information: evidence from managed print services. *Soc. Sci. Electron. Publ. 66*, 392–408. doi: 10.1287/opre.2017.1673

Pérez-Rodríguez, F., Valero, A. (2013). Predictive models: foundation, types, and development. in predictive microbiology in foods. *SpringerBriefs in Food, Health, and Nutrition, vol 5.* Springer, New York, NY. <https://doi.org/10.1007/978-1-4614-5520-2_3>

Rigby, Darrell & Ledingham, Dianne. (2004). CRM done right. *Harvard business review. 82.* 118-22, 124, 126.

Ross, P. T., & Bibler Zaidi, N. L. (2019). Limited by our limitations. *Perspectives on medical education*, *8*(4), 261–264. <https://doi.org/10.1007/s40037-019-00530-x>

Sutton, J., & Austin, Z. (2015). Qualitative Research: Data Collection, Analysis, and Management. *The Canadian journal of hospital pharmacy*, *68*(3), 226–231. https://doi.org/10.4212/cjhp.v68i3.1456

Taylor, S. & Todd, P.A. (1995) Understanding information technology usage: a test of competing models. *Information Systems Research, 6*, 137–176.

Xie, Haoyu. (2021). Research and Case Analysis of Apriori Algorithm Based on Mining Frequent Item-Sets. *Open Journal of Social Sciences. 09.* 458-468. 10.4236/jss.2021.94034.

Yue, H. (2022). Research on E-Commerce Data Standard System in the Era of Digital Economy From the Perspective of Organizational Psychology*. Frontiers in Psychology, 13.*